1-1-2001

Eye movements during recognition of a rotated scene.

Chie Nakatani
University of Massachusetts Amherst

EYE MOVEMENTS DURING RECOGNITION OF A ROTATED SCENE

A Dissertation Presented

by

CHIE NAKATANI

Submitted to the Graduate school of the
University of Massachusetts Amherst in a partial fulfillment
of the requirement for the degree of

DOCTOR OF PHILOSOPHY

February 2001

Psychology

EYE MOVEMENTS DURING RECOGNITION OF A ROTATED SCENE

A Dissertation Presented

by

CHIE NAKATANI

Approved as to style and content by:

Alexander Pollatsek, Chair

Keith Rayner, Member

Neil E. Berthier, Member

Donald L. Fisher, Member

Melinda Novak, Department Head
Department of Psychology

DEDICATION

To my parents.

ACKNOWLEGMENTS

I would like to thank my advisor, Alexander Pollatsek, for his many years of thoughtful, patient guidance and support. I would also like to extend my gratitude to the members of my committee, Kaith Rayner, Neil E. Berthier, and Donald L. Fisher.

I wish to express my appreciation to all individuals who helped me to carry over this project in the USA, the Netherlands and England. Without their implicit and explicit supports, I was not able to complete the project.

Finally, I would like to thank Cees van Leeuwen whose friendship provided me with the encouragement to continue.

ABSTRACT

EYE MOVEMENTS DURING RECOGNITION OF A ROTATED SCENE

FEBRUARY 2001

CHIE NAKATANI, B.A., KWANSEI GAKUIN UNIVERSITY

M.A., KWANSEI GAKUIN UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Alexander Pollatsek

Eye movements during a scene rotation task were measured in two experiments. Two desktop scenes (each consisting of three office objects on a square desktop) were presented consecutively. Participants judged the identity of the two scenes. On *same* trials, the two scenes were either identical or one was a rotated version of the other. On *different* trials, the scene frame was as on the *same* trials, but either the locations or the orientations of some of the objects were changed. Eye movement measures were obtained as real-time indices of information processing.

During the task, the eyes dwell on an object region longer when a scene was rotated further (i.e. *gaze duration* increased) only after the first 900ms of scanning. This result accords to a model in which (a) initial encoding takes place before an alignment process is initiated and (b) alignment is piecemeal and takes place on a gaze-by-gaze basis.

As in previous scene rotation experiments, the slope of a mental rotation function differed between conditions. Response latencies increased more strongly with rotation angle in the orientation-change condition than in the location-change condition. This difference was mainly observed for gaze duration. On the other hand, response times in the Y (vertical)-axis rotation conditions were longer than those in the X (horizontal)- and Z

(line-of-sight)-axis rotation conditions. This difference corresponds to an increase in the *number* (rather than the duration) of gazes in the Y-axis rotation conditions.

Furthermore, when objects switched their locations, the changed object was fixated earlier than an unchanged object. In accordance with this result, it was assumed that the detection of the location-change is handled not only by foveal vision, but also by parafoveal vision. In Experiment 2, the desktop was removed from the scene in half of the conditions. In these conditions location-changed objects no longer were fixated earlier than unchanged objects. Another consequence of removing the desktop was that the eyes need to visit objects more often. This means that desktop frame facilitates the piecemeal alignment process. The results were discussed in terms of viewpoint-dependent models of object recognition.

CONTENTS

Page

ACKNOWLEGEMENT v

ABSTRACT vi

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1. INTRODUCTION 1

Mental Rotation and Scene Recognition 1
Mental Rotation, Scene Recognition, and Eye Movements 5
Essentials for a Process Model for Scene Rotation 11

2. EXPERIMENT 1 14

Introduction 14
Method 19
Results 22
Discussion 36

3. EXPERIMENT 2 43

Introduction 43
Method 46
Results 47
Discussion 58

4. GENERAL DISCUSSION 61

Underlying Processes in the Scene Rotation Task 61
Relation to Models of Object Recognition across Viewpoint Changes 66
Conclusion 69

BIBLIOGRAPHY 87

LIST OF TABLES

TABLE

Page

1. Response times and error rates in Experiment 1 70

2. Response times and error rates in Experiment 1, listed by each stimulus type 71

3. Fixation-0 times, first-pass times, and second-pass times in Experiment 1 73

4. Fixation-0 times, first-pass times, and second-pass times in the same-scene and different-scene conditions 74

5. Response times and error rates in Experiment 2, averaged over all stimulus types 76

6. Response times and error rates in the same-scene and different-scene conditions in Experiment 2 77

7. Fixation-0 times, first-pass times, and second-pass times in Experiment 2, averaged over all stimulus types 80

8. Fixation-0 times, first-pass times, and second-pass times in the same-scene and different-scene conditions in Experiment 2 81

LIST OF FIGURES

FIGURE

Page

1. Scene stimuli 85

2. Same-scene and different-scene conditions 86

CHAPTER I

INTRODUCTION

Mental Rotation and Scene Recognition

In the 1970's and 1980's, Shepard and his collaborators published a series of studies on 'mental rotation' (see Shepard and Cooper, 1982, for an overview), which are among the best-known studies in experimental psychology. Shepard and Metzler (1971) presented 2D images that were projections of 3D figures assembled from cubes (the so-called Shepard-Metzler objects). The objects were rotated various degrees of visual angle in a three-dimensional space with respect to each other. Participants were asked to judge if the 3D objects represented by the two images were the same or different. Shepard and Metzler's main finding was that the response time to process an object increased linearly with the angle of rotation when the object was rotated. They called this linear function the *mental rotation function*, and claimed that the increase in response time was proportional to the angle of rotation because a 3D representation of one of the objects was mentally rotated at a constant angular velocity before comparing it to the other object. This hypothesis is called the *mental rotation hypothesis*.

In the original mental rotation paradigm the stimulus object was a line drawing of connected cubes, but some studies extended the mental rotation paradigm to more naturalistic, complex, or familiar objects. These studies indicated limitations of the mental rotation hypothesis. For example, Jolicoeur (1985) reported that the mental rotation function for a natural object, such as a drawing of a dog, was not linear – the naming time for an upside-down image (i.e. 180-degree rotation) was faster than that of the 120-degree rotated image. Bethell-Fox and Shepard (1988) rotated a complex or

simple matrix pattern in the picture plane, and found that the slope of the mental rotation function for the complex stimuli was greater than that for the simple stimuli. The fact that the more complex pattern was rotated slower than the simpler one suggested that the mental rotation process for a complex object was not holistic, but piecemeal. In a more radical departure from the mental rotation hypothesis, Biederman and Gerhardstein (1993 and 1995) reported that there was no mental rotation needed for familiar objects, such as a telephone: the response time was constant regardless the angle of rotation. They claimed that the recognition of the rotated object is viewpoint-independent when some 3D geometric primitives (i.e., geons) were extracted from the objects. These studies suggest that some revision is necessary if the mental rotation hypothesis is to be applied to more complex and natural viewing situations.

The mental rotation paradigm has recently been applied to scene recognition (Diwadkar and Mcnamara, 1997; Nakatani, Pollatsek, and Johnson, submitted). In these 'scene rotation' studies, a stimulus had several familiar objects that were placed on a background to compose a scene. The scene stimulus employed has, therefore, more complex global-local structure than the Shepard-Metzler object. A major question posed is whether a linear mental rotation function would be observed when a multi-object scene was rotated. Diwadkar and Mcnamara (1997) employed an array of six objects, such as an electric bulb and a mug, on a round table. The object array was rotated around the vertical axis, and the participants were asked to report whether the relative locations of all objects stayed the same after the rotation. The results showed that a mental rotation function was observed in the scene rotation task - the response time increased more or less linearly between the zero-degree and 135-degree rotations.

Nakatani, Pollatsek and Johnson (submitted) reported a similar result with a three-object array. The experiments, however, differed from the Diwadkar and McNamara experiment in two key ways. First, Nakatani et al. examined rotations around four axes, including around the vertical axis and the two other major axes perpendicular to it. Second, two different types of changes, a *location change* and an *orientation change*, were employed. In the location-change conditions, either two or three objects switched their locations on the desktop (e.g., from mug-right and pen-left to mug-left and pen-right), whereas in the orientation-change conditions, either one or all of the individual objects were rotated 90-degrees around their own vertical axes (e.g., from lamp-front-view to lamp-side-view). Thus, in the orientation-change conditions, a rotated object was oriented differently with respect to the desktop, but its location on the desktop was not changed (Diwadkar and McNamara employed only location changes).

Nakatani et al. reported that a mental rotation function was observed not only for a rotation around the vertical axis (i.e., the Y-axis), but also for a rotation around the horizontal axis (i.e., the X-axis) and the line-of-sight axis (i.e., the Z-axis). However, the slopes differed depending on the axis of rotation, as the slope of the mental rotation function was the largest for the Y-axis rotation. Furthermore, the slope varied depending on the type of change made in the scene. Interestingly, the slope of the mental rotation function was steeper in the orientation-change condition than in the location-change condition.

This result is clearly inconsistent with any two-stage model, such as the mental rotation hypothesis, that assumes a completion of alignment of an entire scene followed by a comparison process. If the underlying process for the mental rotation function was a

mental rotation of a complete 3D representation of the scene prior to a global comparison with the representation of the standard scene in memory, the slope (i.e., the rate of alignment) should be the same for the location-change condition and the orientation-change condition (although the absolute times might differ). Instead, the results suggest that a mental rotation process is applied to 'pieces' of the scene stimulus in something like the following fashion. Each piece of the comparison scene is aligned with the standard scene and is compared with the mental representation of the standard scene. The result of the comparison for each piece, a degree of match/mismatch, is accumulated in a buffer for a same/different judgment. The degree of match/mismatch may be accumulated faster in the location-change conditions than in the orientation-change conditions since the slope of mental rotation function was smaller in the location-change conditions than in the orientation-change conditions. In sum, the scene rotation studies showed that (a) a scene with multiple objects in background was processed in a viewpoint-dependent manner when its identity over a rotation was judged, and (b) the alignment process is not holistic.

One of the alternatives to the mental rotation hypothesis is the *viewpoint dependent model* of object recognition (Tarr and Pinker, 1989; Poggio and Edelman, 1990; Edelman and Bülthoff, 1992; Tarr, 1995). The various versions of the viewpoint dependent model claim that the representation of an object does not need to be a complete 3D representation. Rather, a representation that is a collection of local 2D features that are available at each viewpoint is sufficient to explain a mental rotation function. That is, a viewpoint-specific representation of an object is constructed for each view that the participant studied, and recognition of an object is done by matching its

current image to viewpoint-specific representations in memory using an interpolation process. For example, if an object is studied from a "standard" viewpoint so that only a single 2D representation is in memory, and then either an image of the object rotated 15 degrees or 45 degrees is presented, the approximation of the 2D viewpoint-specific representation of the 15-degree rotated image will be closer to the 2D viewpoint-specific representation of the standard than the 2D viewpoint-specific representation of the 45 degree-rotation.

The viewpoint-dependent models may also be able to explain the flat mental rotation function for familiar objects (Biederman and Gerhardstein, 1993). If an object is studied from various viewpoints, multiple viewpoint-specific representations are constructed and any images of the object may be interpolated easily from some viewpoint-dependent representation in memory. If one assumes that familiar objects have been seen many times from various viewpoints, the viewpoint dependent model would predict that a familiar object would be recognized approximately equally well across various angles of rotation and show a relatively flat mental rotation function.

Mental Rotation, Scene Recognition, and Eye Movements

The viewpoint dependent models are by far the most flexible. To some extent, flexibility is a double-edged sword. The models can explain (in principle) how details of the experimental data *might* occur, but are usually not able to not make specific predictions beforehand. Moreover, there is a critical factor missing from the viewpoint dependent models: they do not take eye movements into account. When we inspect a scene, we naturally move our eyes to sample information. In other words, eye movements are an essential component of the dynamic processing in scene recognition. However, a

typical architecture for viewpoint dependent models, such as the RBF network (Edelman and Poggio, 1990; Riesenhuber and Poggio, 1999), is essentially static in its image representations. Thus, such models, in their present format, cannot incorporate the dynamics of the perceptual act. Indeed, the fact that the eyes move 3-4 times a second in most static viewing tasks is a major missing factor in mental rotation studies in general. Although eye movements have been used widely in scene recognition studies (Mackworth, and Morandi, 1967; Antes, 1974; Friedman, 1979; Nelson and Loftus, 1980; Antes and Penland, 1981; Boersema, Zwaga and Adams, 1989; Loftus and Mackworth, 1989; De Graef, De Troy and D'Ydewalle, 1992; Boyce and Pollatsek, 1992; Rayner and Pollatsek, 1992; Blackmore, Brelstaff, Nelson and Troscianko, 1995; Mannen, Ruddock and Wooding, 1997; Henderson and Holligworth, 1998 and 1999), there are few studies that have measured eye movements during a mental rotation task. Even in the few studies that have, eye movements were not always the major focus. Eye movements were often used to monitor the location of fixations, but the duration of the fixations were not systematically examined (e.g., Cave, Pinker, Girogi, Thomas, Heller, Wolfe, and Lin, 1994; Irwin and Carlson-Radvansky, 1996.)

So far, the only serious attempt to incorporate eye movements in the mental rotation paradigm was by Just and Carpenter (1985). They measured eye movements during a mental rotation task with two "cubes". Each cube was a simple line drawing: a hexagon outline contour with a Y junction separating the hexagon into three faces of a cube -- the top, front and right faces. Each face was labeled by a letter or a digit (e.g., "G", "B", and "4"). The standard and comparison cubes were presented simultaneously, and the comparison cube was either a different cube (i.e., some alphanumeric characters

on corresponding faces would be different) or the same cube rotated. Using various test batteries, they selected two groups of participants, high spatial ability participants and low spatial ability participants. The participants were asked to judge whether the two cubes were the same based on the three visible faces, and their eye movements were recorded.

Just and Carpenter hypothesized that the participants would search for a matching pair of characters, determine a trajectory of a "rotation", then confirm the correspondence of the locations and orientations of other characters. The eye movements were analyzed based on this hypothesis. The first sequence of eye fixations that alternated between matching characters on the two cubes was called "the initial rotation". After that, alternation on subsequent matching facets was called "subsequent rotation". Fixations up until the first two matching faces were found (i.e., fixations *before* the initial fixation) were called "search". The remaining fixations were classified as "confirmation". As the angle of rotation increased, (a) the gaze duration of the search stage increased equally for both high-spatial and low-spatial ability groups, (b) the duration of the initial rotation stage increased only for the low-spatial ability group, and (c) the gaze duration of the confirmation stage increased for both groups, but the increase was much larger for the low-spatial ability group. The main focus of the cube rotation study was the difference in processing between the high- and low-spatial ability groups. However, the most general importance of the study is showing the effect of angle of rotation in 'gaze duration'[1]; the

[1] Gaze duration in the cube rotation task was a sum of consecutive fixations made to one of the three faces of a cube. This is somewhat different from conventional gaze duration that is a sum of consecutive fixations made to an object.

gaze duration indeed increased as the angle of rotation increased. In other words, the study allowed for the possibility that 'mental rotation' is applied to a piece of visual information extracted from a gaze.

The results of the cube comparison experiment are interesting, but are hard to generalize to other mental rotation tasks, especially to a scene recognition task with multiple-objects. The cube stimuli were simple line drawings and their "rotation" was not indicated by a contour change -- only by the change of the letters on the faces. Because a viewpoint change (or a rotation of an object) almost always produces changes in the 2D shape projected to the viewer, it is clearly a special case that the 2D shape of the cube stayed the same over rotations of multiples of 90 degrees around principal axes of the cube. Thus, the participants in the cube comparison experiments were not able to use a 2D contour change that would be a natural and an effective cue for the amount of rotation. Participants, therefore, might have relied on reasoning and inference more than on perceptual processes such as comparison among visual features. Furthermore, the operationally defined eye movement stages (search, initial rotation, subsequent rotation and confirmation) make sense only when standard and comparison cube are presented simultaneously. However, when recognizing a scene, the scene before the change is usually not available (unless you have a picture of the scene before the change) and we need to rely on memory. Thus, it is not clear how the results of the cube rotation task could be generalized either to the scene rotation task or an object recognition task.

To understand eye movements in the scene rotation paradigm, one also needs to consider the functional differences between foveal and parafoveal vision. Foveal vision is hightly accurate and based on visual input mainly from the fovea (whose diameter is

about two degrees). Parafoveal vision consists of visual input from the surrounding area of the fovea, up to 10 degrees from the center of retina. Studies have suggested that the difference between foveal and parafoveal vision is not a mere difference in spatial resolution. For example, Pollatsek, Rayner and Collins (1984) examined the level of information extracted from a parafoveal view with a 'preview' technique: While a participant fixated on the center of the screen, a line drawing of a familiar object, such as a cow, was presented parafovealy (5 or 10 degrees from fixation). Participants moved their eyes to the picture and named it as soon as possible. About a half of the time, the preview was switched to another image by the time the eyes arrived. Although participants were not aware of the switch most of the time, naming time increased when the preview and the subsequently fixated view were completely different. Interestingly, when the preview of a cow was changed to an image of another cow, the naming time was faster than that in the different-object preview condition. In other words, the parafoveal information of the visually similar object facilitated the object naming. On the other hand, if the preview-target pair was semantically related, such as baseball bat and ball, there was no benefit from the preview. The results suggest that the level of information extracted from a parafoveal view is higher than contour (there was a considerable contour change between the visually similar preview and target), but lower than semantic category. Similar results were reported by Henderson and Anes (1994) and Henderson (1997).

Foveal and parafoveal vision seem to have different roles and interact in rather complicated ways. Henderson, Pollatsek, and Rayner (1987) combined foveal priming and preview. The participant sequentially fixated on two objects (horizontally separated

by 5 or 10 degrees), and named the second object. The two objects were semantically related (prime and target, such as cat and dog) on a third of the trials. At the same time, the preview of the second object was manipulated. On 50% of the trials, the preview of the second object was not available; a blob or a placeholder was presented while the eyes were fixating on the first object. Henderson et al. reported that the naming time was faster when a preview was available (i.e., preview benefit). Also, the effect of foveal priming was smaller in the 5-degree eccentricity condition than in 10-degree eccentricity condition. However the foveal priming effect was not additive with the preview benefit; the priming effect was smaller when the preview was available. These results suggest that the foveal prime information was sent to a semantic network and it took some time for activation to spread to its semantic neighbors. When the degree of separation was small (i.e., the saccade duration was short), the processing of the second object did not have enough time to obtain a full benefit of the foveal priming. It is not clear why the presence and absence of the preview affected the size of the foveal priming benefit; however, a similar non-additivity was also reported by Henderson (1992).

In sum, these object identification and eye movement studies suggest that (a) the information within a fixation is processed and used differently depending on its eccentricity, (b) foveal information is sent to a higher-level information processing system, such as activating a semantic network. On the other hand, (c) parafoveal information facilitates the identification of objects that are later fixated. Furthermore, (d) the level of information carried over a saccade seems to be 'intermediate' between physical features, such as contours and semantic features (such as category information).

Essentials for a Process Model for Scene Rotation

Given that the scene rotation task is a visual task that involves eye movements, I shall assume that a working model for the scene rotation task shares many of the same characteristics with the models proposed to explain object identification and eye movement studies. In the present scene rotation task, the participants tried to detect either a location change or an orientation change. I assume that on each fixation, the presence or absence of change is mainly determined from the foveal information, and the result of the processing on that fixation, such as likelihood of presence/absence of a change, is sent to a higher-level process for the same/different judgment. At the same time, the next fixation location is determined from parafoveal information and the processing of the object at the next fixation location starts with processing the parafoveal information obtained about that object. One assumption will be that the alignment process is applied only to the foveal information. Once an alignment is applied to the foveal information, the result is sent to an accumulator. The sequence of fixations and saccades is repeated until enough evidence for the same/different judgment is collected. In other words, the model assumes piecemeal alignment in the framework of a serial self-terminating scan.

As indicated earlier, the idea of piecemeal alignment was advocated by mental rotation studies that used complex stimuli, such as a checkerboard pattern (Folk and Luce, 1987; Bethel-Fox and Shepard, 1988). These studies reported that the slope of the mental rotation function increased when the stimulus was complex. Bethel-Fox and Shepard (1988) wrote, "These patterns naturally yielded an objective measure of complexity, namely, the number of perceptually distinct pieces, that seemed especially

relevant in view of suggestions by several researchers that multipart objects are mentally transformed piece by piece. (p. 13)". Scene stimuli are similar to the above description of "multipart objects". Also, as argued above, the results from our previous study (Nakatani et al., submitted) supported a piecemeal alignment process in the scene rotation task.

Thus, in present study, it was assumed that the alignment process is applied to an object (or more precisely, to an object region) captured by a gaze. With eye movement measures, this assumption can be tested directly. If the assumption is correct, the greater the angle by which the scene is rotated, the longer the eyes will stay on an object. Bethel-Fox and Shepard also claimed that the mental representation made by piecemeal alignment was eventually integrated into a single representation. Bethel-Fox and Shepard assumed that such integration occurs after completing multiple trials. However, similar integration might also progress within a trial. With eye movement measures, it is possible to see the temporal development directly. If the integration of piecemeal representations progresses within a trial, the gaze duration at the beginning of a scan might be longer than that at the end of the scan.

The roles of foveal and parafoveal vision in scene rotation were also explored. Although the visual information from the foveal view is assumed to be of primary use for the detection of a change in the scene, parafoveal vision may facilitate the same/different judgment to some extent. That is, it is likely that the primary basis for a same/different judgment is aligning and matching the foveal information to the corresponding part of the memory representation of the standard stimulus. However, parafoveal information may also play a role. In particular, a preliminary indication from parafoveal vision that there

is a change in some part of the display may help guide the eyes to a changed object. If so, the eyes may fixate on a changed object earlier than an unchanged object.

When eye movements are taken into account, process models become more specific -- and complex -- than non-eye movement models. To appreciate the power of how eye movement measures make a process model more specific (without getting lost in the complexity), I will start from some simple assumptions. Such assumptions can be amended later to obtain more realistic models. For this purpose, a couple of simple eye movement process models are proposed and tested in Experiment 1.

CHAPTER 2

EXPERIMENT 1

Introduction

In Experiment 1, eye movements during the scene rotation task were recorded as well as the response times and error rates. The scene rotation task was the same as that in Nakatani et al. (submitted), discussed earlier.

Consider two simple process models for how the task might be accomplished. Model 1 assumes that a piece of a comparison scene is sampled via foveal vision and then aligned to the mental representation of the standard scene. It is assumed that the participants would have a good mental representation of the standard scene because the same standard scene was presented prior to the comparison scene on each trial of a trial block, for as long as the participant wished. The dimensionality of the representation, 2D or 3D, is not a critical issue here, but to keep the model simple, the initial representation of the comparison scene is assumed to be 2D or "view-based" (Tarr and Bülthoff, 1999). The model also assumes that some crude 2D configurations of the scene, such as the positions of the objects, are computed first from the comparison stimulus to prepare for eye movements. Once the 2D configuration of the comparison scene is extracted, the first saccade to one of the objects is planned and executed. As soon as the eyes are directed to an object (or more precisely an object region), the visual information of the object region is processed. This information (largely foveal) is just a piece of an entire comparison scene. Thus, it needs to be 'moved' to the corresponding part of the mental representation of the standard scene prior to alignment in order to be able to judge whether there was change in that portion of the desktop or not. Any available features

are used to 'move' the piece to a corresponding part of the mental representation. It should be noted that the 'moving' process is *not* an alignment. It is merely pinning the 2D piece on a part of the 2D mental representation of the standard scene where the piece is aligned. Thus, the 'moving' process is making a rough correspondence prior to an alignment. Once the correspondence is established, an alignment is applied between the 2D piece from the comparison stimulus and the mental representation of the standard stimulus. The alignment process will take longer for the larger angle of rotation. After alignment, a comparison process is run to determine whether the foveal information matches the mental representation of the corresponding portion of the standard scene. In this model, the eyes are assumed to stay on the object until both the alignment and comparison processes are completed. Thus, the eyes will stay longer on an object when the angle of rotation of the scene is increased.

The comparison process outputs a degree of match/mismatch between the foveal input and the mental representation. For the sake of simplicity, Model 1 assumes that the comparison results are always sufficient to make a correct same/different judgment on the gazed input. If the object is judged as 'changed' either in location or orientation, the 'different' key is pressed. In other words, a 'different' response is assumed to be based on a different judgment on the input from a single gaze. In contrast, all three objects need to be gazed for a 'same' response. That is, if the object is judged as 'unchanged', the eyes move to the next object and repeat the alignment and comparison process, and if the last object is judged as 'unchanged', the 'same' key is pressed. In other words, the 'same' response is made only after all objects were visited and judged 'unchanged'. Thus, the algorithm of Model 1 is; Go to an object, increment the 'visited' counter, and

perform alignment and comparison. If a gazed object is 'changed (=1)', press the 'different' key – otherwise if the object is 'unchanged (= 0)', check the 'visited' counter. If the counter shows a value less than a maximum (the maximum is the number of all objects), move to the next object, increment the 'visited' counter and repeat alignment and comparison. If the 'visited' counter shows the maximum value and currently gazed object is 'unchanged', press the 'same' key. Thus, Model 1 predicts that the number of 'visits' will be the largest (= the number of all objects) in the same-scene condition, and there will be about the same number of visits made for the location-change and orientation-change conditions. Also, because each alignment is assumed to be conclusive to make a judgment on presence/absence of change on an object, revisits to objects are not necessary.

Model 1 might be too restrictive because it assumes that a single "yes-no" same/different judgment is made at each piecemeal alignment. In Model 2, the result of each alignment is not necessarily sufficient to trigger a 'different' key press. Instead, degree of mismatch is sent to an accumulator. The accumulator has a response threshold, and when the pooled mismatch value exceeds the threshold, the 'different' key is pressed. When the threshold is not exceeded after some number of visits, the 'same' key is pressed. In other words, a 'same' response deadline is set in terms of the number of gazes. Thus, the algorithm for Model 2 is; Go to an object, increment the 'visited' counter, and perform alignment and comparison. Send the comparison result (between 0 = unchanged and 1 = changed) to an accumulator. Check the accumulated value. If the value is more than a threshold, press the 'different' key. If the accumulated value is less than the threshold, check the 'visited' counter. If the counter is less than a maximum (an

arbitrary number within a reasonable range, say up to 10), go to the next object and repeat alignment and comparison. If the counter shows the maximum value, press the 'same' key. The response rules in Model 2 are more complex than those in Model 1 -- the 'different' response is regulated by the threshold of the accumulator, and the 'same' response is controlled by the maximum value of the 'visited' counter. It should be noted that the maximum value is not necessarily the number of the objects. The maximum number could be large, such as 10, if a participant uses a careful double-checking strategy.

Most importantly, in contrast with Model 1, Model 2 predicts substantial differences between the location-change and orientation-change conditions in terms of the number of "visits" to objects. Because the degree of mismatch value will be larger in the location-change condition than in the orientation-change condition, the mismatch value will be accumulated faster in the location-change condition than in the orientation-change condition. For example, in the location-change condition, the foveal information of an object (e.g., mug) would be aligned with the representation of a different object (e.g., briefcase), whereas in the orientation-change condition, the foveal information (e.g., briefcase) is matched against the same object facing a different direction (e.g., 90-degree rotated briefcase). Thus, substantially more visits should be necessary in the orientation-change condition than in the location-change condition to achieve sufficient "difference" to reach threshold. To summarize, Model 2 predicts the largest number of visits in the same-scene condition (because a scan is not terminated until the maximum number of visits are made in the same-scene condition), the second largest number of visits in the orientation-change condition, and the smallest number of visits to the

location-change condition. Moreover, Model 2 allows revisits, because the visit-and-align routine will be repeated until either the 'different' threshold was exceeded or the maximum number of visits was made.

Eye Movement Indices. To test these models, *gaze duration*, *first-pass time* and *second-pass time* were chosen as the principal eye movement indices. Gaze duration is the sum of fixations made to an object before the eyes leave the object. For example, if a series of fixations were made to Object1, Object1, Object2, Object2, Object1, Object3, Object3, and Object1, the sequence of the fixations were grouped based on the fixated objects, such as (Object1 - Object1) - (Object2 - Object2) - (Object1) - (Object3 - Object3) - (Object1). The gaze duration is the sum of the fixation durations within a pair of parentheses. The first-pass time on the scene is the sum of the fixation durations before the first regression back to a previously fixated object, and the second-pass time on the scene is the sum of the fixation durations after the first regression back to a previously fixated object[2]. In the example above, the eyes regressed at the fifth fixation. Thus, the first pass time is the sum of the durations of the first four fixations and the second-pass time is the sum of the fixation durations of the last four fixations. The gaze duration, first-pass time and the second-pass time measures were adopted from text processing studies where eye movement records have been used extensively (See Rayner, 1998, for a review). In the text processing studies, the gaze duration is defined as the

[2] The first-pass time in this article is, however, slightly different from that in the text comprehension studies. In the text comprehension, the first-time fixation to a word is included to the first-pass time even after the first regression (in this paper, all fixations after the regression was excluded from a first-pass time). In an alternative analysis, the data in this paper were analyzed in the same way as in the text comprehension studies. However, the current method – dichotomy at the first regression – was preferred as it provided more clear-cut results.

sum of the durations of consecutive fixations made to a word. In the scene rotation task, an object was taken as a unit equivalent to a word in reading. If the eyes start moving clockwise from the first fixated object (there are only two directions of scan, clockwise or counterclockwise, because there were only three objects), the first counter-clockwise search is taken as the first regression.

In addition to the fixation-duration based indices, the number of gazes was used to test whether parafoveal information is used to help guide eye movements (and hence processing). If a location-change or orientation-change is signaled by the information in parafoveal vision, a changed object will be fixated earlier than an unchanged object.

Method

Participants. Twenty-one undergraduate students and graduate students of the University of Massachusetts, two men and 19 women, participated in the experiment. All of the participants had normal or corrected-to-normal vision. They received $8.00 or experimental credits in psychology courses for their participation.

Stimuli and Design. The stimuli were computer-generated images of office objects on a desktop. There were four different scenes. Each scene had three office objects on a square desktop: Scene 1 – briefcase, mug, and calculator; Scene 2 – stapler, keyboard, and monitor; Scene 3 – pen, telephone, and tape dispenser; Scene 4 – desk lamp, document box and index card holder. The desktop and the objects were made by 3D graphic models with the Infini-D 2.5 software package. Each scene was rendered as an 800 by 600 pixel image with naturalistic colors and shading. The objects were carefully placed on the desktop to prevent any occlusion among the objects in any of the views.

The standard viewpoint was from the front of the desktop, five degrees above the gravitational horizontal plane of the desktop (see Figure 1). There were seven alternative viewpoints created by rotating the (hypothetical) camera around the desktop, starting from the standard viewpoint. There were three axes of rotation. The X-axis went from left to right, so that X-axis rotations were equivalent to bringing the viewer over the desk top, and the viewer got close to a "bird's-eye" view of the desktop at the 70-degree rotation. The Y-axis was the vertical axis and the Y-axis rotations were clockwise in the horizontal plane, so that the view was as if one were walking around the desk to the left. The Z-axis went straight out from the viewer, so that a Z-axis rotation was a counterclockwise rotation of the camera in the picture plane. For each axis of rotation, there were two levels of rotation: one in which the camera was rotated 35 degrees and one in which it was rotated 70 degrees. In addition to the rotations around a single axis, a double axis rotation was included. The double axis rotation was a composite rotation around two axes, 70 degrees around the Y-axis plus 70 degrees around the X-axis. The eight viewpoints, the standard (or 000), X35, X70, Y35, Y70, Z35, Z70 and double-axis (Y70-X70), were the same as those used in Experiment 2 and 3 of Nakatani et al.

For each of the eight comparison viewpoints, there were three different stimulus types (see Figure 2). In the *same-scene* stimuli, there was no change in either location or orientation of any of the three objects relative to the desktop. There were two types of *different-scene* stimuli. In the *location-change* condition, the locations of two objects were switched. In the *orientation-change* condition, an object was rotated +90 or –90 degrees around its own vertical axis, but there was no location change. Furthermore, there were two levels of change in both the location-change and orientation-change

conditions. In the location-2 condition, two of the three objects switched their locations, whereas in the location-3 condition, all three objects switched locations. Analogously, in the orientation-1 condition, only one object was rotated +90 or –90 degrees around its own vertical axis, while all objects were rotated +90 or –90 degrees in the orientation-3 condition. All these changes were approximately counterbalanced across particular objects and viewpoint changes.

Procedure. The task was a judgment of whether the object arrangements in the standard and comparison scenes were the same or different. The standard scene, which was always in the standard viewpoint, was presented before each trial, and participants could view it as long as they needed. When they pressed the "ready" key (the space bar), a pattern mask was presented for 300 ms, and then a fixation point was presented. As soon as the participant fixated on the fixation point, the comparison scene appeared and remained until the participant responded "same" or "different". The '/' key was assigned to the "same" response and the 'z' key was assigned to the "different" response. The participants were asked to respond as accurately as possible.

The participants received 32 practice trials with feedback, and then completed the 256 trials of the experimental session without feedback. The practice trials used a different scene from those used in the experimental session, but were otherwise similar. The 256 experimental trials were divided into four blocks of 64 trials. In each block, all trials were with the same standard scene. In each block, half the trials were the same-scene trials and half were the different-scene trials. For the 32 different-scene trials, two location-change trials and two orientation-change trials were presented for each of the eight viewpoints. To make the number of the same-scene and different-scene trials

equal, the same-scene stimulus at each of the eight viewpoints was repeated four times. The order of the 64 trials within a block was randomized separately for each participant, and the order of the standard scenes was counterbalanced over participants.

The eye movements during the scene rotation task were measured by the SMI EyeLink system. The system consists of a lightweight helmet that has an infrared sensor for head movements and two CCD cameras for eye tracking, two Compaq Deskpro PC compatible computers and a 17-inch Viewsonic 17PS monitor. After the practice session, the participants donned the helmet and were seated in front of the display. Head movements were monitored and canceled out by the IR sensor system, thus no head support was used. The display-eye distance was approximately 80 cm. Eye position was only recorded from the right eye. The camera position and image level were adjusted, and then the nine-point grid calibration was begun. In general, the calibration was conducted at the beginning of each block. An automatic drift correction was made around the center point before each trial, and a re-calibration was inserted when the current measurement fell out of margin of the drift correction. The refresh rate of the display was 75 Hz, and the sampling rate of the eye position was 250 Hz. The sample was filtered and fixation durations and fixation locations were extracted along with other saccade-related indices. The filter was a part of the EyeLink System which computes the saccade-related indices based on acceleration.

Results

General Modes of Analysis. There were five types of trial: *same-scene* trials and four types of *different-scene* trials (*location-2, location-3, orientation-1,* and *orientation-3*). The design was not completely factorial because there was only a single zero degree

rotation condition. As a result, two different types of analyses were used to assess the rotation effects for each axis and the differences in rotation effects across axes. In the first, the slope of rotation for each axis of rotation was assessed by the difference between the 70-degree rotation condition for that axis and the no rotation condition, divided by 70 degrees. (For three equally spaced values, the linear trend test is a contrast between the extreme values.) Because the no rotation condition was a common baseline, differences among the axes in the slope of rotation were assessed by comparing the 70-degree rotation conditions. In the second type of analysis, rotation effects were assessed by comparing the 35- and 70-degree rotation conditions; here, a factorial analysis was used.

To simplify exposition, the mental rotation function of the scene rotation task will be initially described averaged over all five types of comparison trials and then the data will be examined separately for each type of comparison. Moreover, the results in the single axis rotation conditions (i.e., the X-, Y- and Z-axis rotation conditions) are reported first, then, the results in the double axis rotation condition (Y70-X70) are discussed.

Details of Computation of the Eye Movement Indices. All the eye movement indices were computed from correctly answered trials. Thus, 132 error trials, which were 2.47% of all trials, were excluded from the analyses. Some of the error trials were analyzed separately and are discussed later. Prior to the computation of the eye movement indices, fixations that fell outside the borders of the screen image (600 by 800 pixels) were excluded. The total number of excluded fixations was 700, which was 2.26

% of the total number of fixations. Ten trials in which the recording of the eye movements was disturbed were also excluded from the data analysis.

At the beginning of a trial, the participant was fixating on the fixation circle placed at the center of the display. The fixation circle was replaced by the comparison stimulus, which began the epoch in which eye movements were analyzed. The time from the onset of the comparison stimulus to the start of the first saccade was analyzed separately from the rest of fixations because the fixation - *fixation-0* - was not on any of the objects. Thus, the fixation after the first saccade from the fixation circle was taken as the first fixation. The fixation durations of the first fixation to the last fixation before the key-press were analyzed in various ways. (Averaged over all types and viewpoints, 5.71 fixations – excluding fixation-0 -- were made per trial.) The fixation duration of the last fixation was judged as terminated by the offset of the comparison stimulus even though it usually lasted longer. It should be also noted that saccade durations were not included in the sum of the fixation durations, such as first-pass time and second-pass time.

In formulating eye movement measures, one has to decide which object is being fixated on each fixation. This, however, is not straightforward because visual information is extracted from a wider visual field (more than 15 degrees) in picture viewing than in reading (less than 5 degrees). Thus, a viewer's fixation point does not have to be as close to an object to identify it as it does with a word. In this study, the closest object from the fixation location was defined as the fixated object. The distance measure used was the Euclidean distance between the fixation location and the center of the object, where the center of the object was estimated by the experimenter. The average distance between the fixation location and the center of the closest object was 45

pixels. As the objects were approximately 100 by 100 pixels on average, most of the fixations were indeed made on the objects.

On 25 trials (0.4% of all trials), the participants did not make an eye movement -- they terminated the trial by a button-press without moving the eyes from the fixation circle to any of the objects. Interestingly, all the no-eye-movement trials were correctly answered. However, these were also excluded from the main analyses because they seemed qualitatively different and will be discussed separately.

In next section, the response times and error rates are presented first to show that the mental rotation function that was reported in the previous scene rotation experiment of Nakatani et al. was replicated. Second, the two process models were examined using the eye movement data.

Response Time and Error Rate. Averaged over all stimulus types, the response times increased when the angle of rotation increased (See Table 1). The slopes of the X-, Y- and Z-axis rotation functions were 2.64, 4.90, and 4.90 ms/deg., respectively. All three slopes were significantly greater than zero, $ts(20) > 2.80$, $ps < .05$. The slopes of the Y- and Z-axis rotation conditions were both steeper than that the X- axis rotation condition, $ts(20) > 2.49$, $ps < .05$. The error rates increased slightly as the angle of rotation increased. However, the increase was significant only for the Y axis rotation, $t(20) = 1.67$, $p > .1$, $t(20) = 2.71$, $p < .05$, and $t < 1$, for the X- Y- and Z-axis, respectively. In sum, viewpoint-dependent processing was once again observed in the scene rotation task.

Table 2 shows the response times and error rates in the same-scene and four different-scene conditions. The response times in the same-scene conditions were slightly longer than those in the different-scene conditions, but the difference was not

significant, $t(20) = 1.19$, $p > .1$. However, the error rates in the same-scene conditions were lower than those in the different-scene condition, $t(20) = 3.80$, $p < .001$. Thus, a significant fraction of the errors were likely due to missing a change when it was present.

The response times in the same-scene condition increased significantly for the 70-degree difference in all axes of rotation, $ts(20) > 3.13$, $ps < .01$. However, in the location-change conditions, the slopes of the rotation functions were not significantly greater than zero, $ts(20) < 1.41$, $ps > .1$, except for the Z-axis rotation in the location-3 condition, $t(20) = 4.67$, $p < .001$. The slope was even negative for the X-axis rotation of the location-2 condition, $t(20) = 2.33$, $p < .05$. In the orientation-change conditions, response times increased significantly as the angle of rotation increased, $ts(20) > 2.51$, $ps < .05$, except for the Y-axis rotation in the orientation-1 condition, $t(20) = 1.07$, $p > .1$, and the X-axis rotation in orientation-3, $t < 1$. The response times in the orientation-change conditions were slightly longer than those in the location-change conditions, $F(1, 20) = 3.27$, $p < .1$. The difference in rotation effect between location-change and orientation-change is also indicated by an interaction of location- vs. orientation-change with the angle of rotation, $F(1, 20) = 12.52$, $p < .01$. The orientation-1 condition also showed by far the highest error rates among the four different-scene conditions, $F(1, 20) = 27.17$, $p < .001$, for the interaction between the location- vs. orientation-changes and the size of the change (i.e., location-2 and orientation-1 vs. location-3 and orientation-3 conditions). Overall, the response time and error data showed a very similar pattern to that of our previous scene rotation experiments (Nakatani et al., submitted), except for the location-change conditions.

Testing the Process Models. The process models made different predictions regarding the number of gazes. Thus, the number of gazes was analyzed first. Model 1 assumes that a result of an alignment and comparison is evaluated at each gaze (i.e., a 'visit' to an object). The model thus predicts that (a) more gazes are made in the same-scene condition than in the location- or orientation-change conditions, and (b) the number of gazes is about the same in the location-change condition and the orientation-change condition. In Model 2, the results of alignment and comparison, degree of mismatch, are not evaluated immediately, but are accumulated as a scan over the scene progresses. A comparison scene is judged 'different' only after the pooled mismatch value exceeds a threshold. If the threshold is not exceeded after some number of gazes, the scene is judged 'same'. The model presumes that the rate of accumulation of mismatch is slower in the orientation-change condition than in the location-change condition. Thus, Model 2 predicts the largest number of gazes in the same-scene, the second largest number of gazes in the orientation-change condition, and the smallest number of gazes in the location-change condition.

The data showed that the number of gazes in the same-scene condition was indeed the largest, and the number of gazes in the location-change and orientation-change conditions were about the same. (Only the orientation-3 and location-3 conditions were used for the comparison to have an equal number of changed objects.) Averaged over the seven viewpoints (000, X35, X70, Y35, Y70, Z35 and Z70), the number of gazes were 4.43, 3.78, and 3.63 in the same-scene, orientation-3, and location-3 conditions, respectively, $t(20) = 4.75$, $p < .001$, for the same-scene vs. orientation-3, $t(20) = 4.80$, $p < .001$ for the same-scene vs. location-3, $t < 1$, for the orientation-3 vs. location-3. The

qualitative pattern of data is thus closer to what Model 1 predicts than what Model 2 predicts. However, there is a large discrepancy between the values predicted by Model 1 and the observed data. As Model 1 assumes that a same/different judgment is made at each gaze, the expected number of gazes predicted by Model 1 is, 3, 1, and 1 for the same-scene, location-3, and orientation-3 conditions, respectively. This disagreement is discussed after the presentation of other eye movement data.

Further analyses were conducted on the number of gazes. The number of gazes was dependent on the angle of rotation. Averaged over the same- and different-scene conditions, the number of gazes increased with the angle of rotation. When the scene was rotated 70 degrees, the number of gazes increased from 3.77 to 4.04 (+0.27 gazes) for X-axis rotations, $t(16) = 2.06$, $p = 0.05$, from 3.77 to 4.43 (+0.66 gazes) for Y-axis rotations, $t(16) = 7.67$, $p < .001$, and from 3.77 to 4.12 (+0.35 gazes) for Z-axis rotations, $t(16) = 3.54$, $p < .01$. It should be noted, however, the number of gazes was not the only cause of the rotation effect in response times – as described later, gaze duration also increased for the larger angle of rotation. Furthermore, the number of the gazes varied among the axes of rotations; the number of the gazes in the Y-axis rotation condition was larger than that of the X- and Z- axis rotation conditions $t(16) = 3.04$, $p < .001$, $t(16) = 2.51$, $p < .05$, for Y vs. X, and Y vs. Z, respectively, but there was no significant difference between the X- and Z-axis rotation conditions, $t < 1$. More detailed analyses on the number of gazes are described in later sections.

Late Onset of Alignment. The analyses of first-pass and second-pass times suggested that there was a complex pattern in the underlying processes. Table 3 shows the first-pass time, second-pass time and fixation-0 time (fixation-0 is discussed later)

averaged over the same-scene and different-scene conditions. Interestingly, the first-pass time was hardly affected by the angle of rotation, but there was a substantial increase in second-pass time as the angle of rotation increased. The slope in the second-pass time was larger than that on the first-pass time, $F(1, 20) = 17.98$, $p < .001$. The first-pass time slopes were not significantly larger than zero except for the Z- axis rotation, $t < 1$, $p > .1$, $t(20) = 1.07$, $p > .1$, $t(20) = 2.11$, $p < .05$, for the X-, Y- and Z-axis respectively. The results suggest that alignment did not take place during the first pass, except possibly for the Z-axis rotation condition. This is puzzling because the first-pass time was about 1.5 times longer than the second-pass time averaged over the seven viewpoints, which suggests active information processing. Thus, the initial scanning process summarized in the first-pass time appears to be some kind of 'encoding' necessary for the subsequent alignment process occurring during the second pass.

The gaze durations in the first and second passes were also analyzed and, averaged over seven viewpoints, the mean gaze duration was about 30 ms longer in the first-pass (391 ms) than in the second-pass (360 ms), $t(20) = 3.38$, $p < .01$. Comparing the 0- and 70-degree conditions, the mean gaze duration did not increase with increasing rotation angle during the first pass except for the Z-axis rotation condition, $t < 1$, $t < 1$, $t(20) = 2.60$, $p < .02$, for X-, Y- and Z- axis rotation, respectively. In contrast, during the second-pass, the mean gaze duration increased around 67 ms between the 0-degree and 70-degree for all axes, $t(20) = 2.62$, $p < .001$, $t(20) = 4.42$, $p < .001$ $t(20) = 3.09$, $p < .01$, for X-, Y- and Z- axis rotation, respectively. Thus, the gaze duration data thus support the idea that a piecemeal alignment is applied at each gaze.

One of the major findings in Nakatani et al. (submitted) was a different slope for the mental rotation function in the orientation-change and location-change conditions; the slope was larger in the orientation-change condition than in the location-change condition. The eye movement data in this experiment showed that the gaze duration increased more in the orientation-change condition than that in the location-change condition. Averaged over all axes of rotation, the gaze duration during second-pass increased from 313 ms for the no rotation condition to 465 ms for the 70-degree rotation in the orientation-change condition. In contrast, in the location-change condition, the second-pass gaze duration did not increase for the 70-degree rotation condition, but decreased from 390 ms to 369 ms. The rate of increase was tested by a 2x2x3 ANOVA (location- vs. orientation-change, large- vs. small-changes, and axes of rotation. The main effect of the location- vs. orientation-change was significant, $F(1, 20) = 5.62$, $p < .05$[3]. At the same time, the rate of increase in the number of gazes during the second-pass is slightly larger in the orientation-change condition, and the difference was marginally significant, $F(1, 20) = 4.33$, $p = .051$[4]. The results thus indicated that the slope of the mental rotation function was larger in the orientation-change than that in the location-change condition because both the duration and number of gazes during the

[3] A 2-way interaction between the location- vs. orientation-change and large- vs. small-changes, and the 3-way interaction were not significant, $Fs < 1$. Thus the second-pass gaze duration in the orientation-1 condition did not increased more than that in the orienation-3 condition when the angle of rotation was increased.

[4] The same 2x2x3 ANOVA as in the second-pass gaze duration was used. In the number of gazes, the main effect of the size of change was significant, $F(1, 20) = 4.39$, $p < .05$. A 2-way interaction between the location- vs. orientation-change and large- vs. small-changes was not significant, $F < 1$, but, the 3-way interaction was, $F(2, 40) = 8.79$, $p < .001$.

second-pass increased more for the greater angle of rotation in the orientation-change condition than in the location-change condition.

Fixation-0 Time. As mentioned earlier, the latency from the onset of the comparison stimulus to the start of the first saccade was excluded from the above analyses of fixation durations because one doesn't know what object or objects are being processed when the eyes are initially fixating in the center. This latency, fixation-0 time, however, is also a part of the response time and it would reflect the very beginning of processing. Thus, the fixation-0 time is described separately in this section. The fixation-0 times did not increase with increasing angles of rotation. There seemed to be two groups: the fixation-0 times in the conditions in which there was no rotation in depth (the 000, Z35 and Z70 conditions) were around 276 ms, and those in the other conditions (the X35, X70, Y35, and Y70 conditions) were around 214 ms. A one-way F-test for the Z-axis rotation conditions showed no significant difference between the Z35 and Z70 (picture plane rotation) conditions and the 000 (no-rotation) condition, $F < 1$. On the other hand, the fixation-0 time in the 000 condition was longer those that in the X- and Y-axis rotation conditions, $F(2, 40) = 19.76$, $p < .001$, $F(2, 40) = 39.52$, $p < .001$, respectively. Post-hoc pair-wise contrasts showed that fixation-0 time in the 000 was longer than those in all of the X35, X70, Y35 and Y70 conditions, $ts(20) > 4.19$, $ps < .05$, adjusted by using the Bonferroni method.

This two-group pattern was common to the various comparison types: In the same-scene condition, the fixation-0 times in the 000, Z35 and Z70 conditions were longer than those in the rest of conditions (See Table 4). There was no significant difference among the 000, Z35 and Z70 conditions, $F(2, 40) = 1.46$, $p > .1$. The shortest

fixation-0 time of the three was 269 ms in the Z70 condition. When the fixation-0 time in the Z70 condition was compared with that of the X35, X70, Y35 and Y70 conditions, the difference was significant for all comparison, $ts(20) > 2.95$, $ps < .02$, adjusted by the Bonferroni method. There were some variations among the four different-scene conditions. However, the fixation-0 times in the 000, Z35 and Z70 conditions were still longer than those in the X35, X70, Y35 and Y70 viewpoints, and there was no significant difference among the 000, Z35 and Z70 conditions, $F < 1$. The shortest fixation-0 time of the three, 275 ms in the Z35 condition, was still longer than those in the X35, X70, Y35 and Y70 conditions, $ts(20) > 3.29$, $ps < .02$, adjusted by using the Bonferroni method. The results showed that the eyes stayed longer at the center of the display when the comparison stimuli were roughly the same as the standard stimulus in 2D. The results suggest some kind of 2D configuration was extracted and used before the eyes start moving.

Use of Information from Parafoveal Vision. Another point of interest is the function of parafoveal vision in the scene rotation task. Object identification studies (Pollatsek et al, 1984; Henderson and Anes, 1994; Henderson, 1997) suggested that parafoveal vision facilitates foveal processing. In the scene rotation task, parafoveal vision might facilitate foveal process in several ways. One way is that of making a preliminary judgment about where is a change and then guiding the eyes to the possibly changed object. If so, a changed object would be fixated earlier than an unchanged objects. To test the prediction, the number of gazes was counted in the location-2 and orientation-1 conditions before a changed object was fixated. To remedy the different number of changed object in these conditions, the number of gazes in the same-scene

condition until that object was fixated was used as baseline. For example, in the orientation-1 condition, if the mug was changed in the X35 viewpoint of Scene1, the number of gazes until the mug in the X35 same-scene condition was fixated was used as the baseline. In the location-2 condition, the baseline was the number of gazes until *either* of the 'changed' objects was fixated in the corresponding same-scene trial. Averaged over all viewpoints, it took 1.31 gazes to reach the changed object in the location-2 condition. This was smaller than its baseline, 1.48 gazes, $t(20) = 5.10$, $p < .001$. However, in the orientation-1 condition, it took 1.91 gazes, which was actually significantly larger than the baseline, 1.76 gazes, $t(20) = 3.56$, $p < .01$. This is counterintuitive and it is not clear why the latter effect was observed. In sum, the results showed that parafoveal vision helped guiding the eyes to the changed object in the location-change conditions, but it was not the case in the orientation-change conditions.

Double Axis Rotation Condition. In the double axis rotation condition, the comparison stimuli were rotated 70 degrees around the Y-axis, then rotated 70 degrees around the X-axis (See Figure 1). However, the rotation time in the double axis rotation condition was much shorter than the sum of the rotation time of the Y70 and X70 conditions (See Table 1). The response time in the Y70-X70 condition was about the same as the X70 condition, $t < 1$, but was shorter than that in the Y70 condition, $t(20)= 3.81$, $p < .002$. This pattern is common for all types of trials (See Table 2); the response time in the double-axis rotation condition was about the same as that in the X70 condition, $ts(20) < 1.67$, but was shorter than that in the Y70 condition, $ts(20) > 2.53$, $ps < .05$, except for the location-3 condition, $t < 1$. When averaged over the types, the error rate in the double-axis rotation was about the same as that in the X70 condition, $t < 1$, but

smaller than that in the Y70 condition, $t(20) = 3.81$, $p < .001$. In sum, the double-axis rotation condition showed a similar pattern to the X70 condition for both response time and error rate.

The eye movement indices in the double axis rotation condition also showed a similar pattern to that in the X70 condition rather than that in the Y70 condition. The first-pass time of the double-axis rotation condition was longer than those in both X70 and Y70 conditions, $ts(20) > 2.29$, $ps < .05$. In the second-pass time, the similarity between the double-axis condition and the X70 condition was clear; the second-pass time in the two conditions were about the same, $t(20) = 1.96$, $p > .1$, whereas the second-pass time in the Y70 condition was 270 ms longer than that in the double-axis condition, $t(20)$ 5.65, $p < .001$. The fixation-0 time in the double axis rotation condition was about the same as that in the X70 and Y70 conditions, $t(20) = 1.89$, $p < .1$ for the double-axis vs. X70 conditions, $t < 1$, for double-axis vs. Y70 conditions. The fixation-0 time of the double axis rotation condition (204 ms) was closer to the other rotation in depth conditions (whose average was 216 ms) rather than the group of the 000, Z35, and Z70 conditions (whose average was 276 ms).

In sum, the results of the double axis rotation condition were similar to those of the X70 rotation. Parsons (1987) concluded that the object in the multi-axis rotation conditions was rotated around a unique axis that makes the shortest pass rotation by examining the mental rotation functions of Shepard-Metzler objects in various multi-axis rotation conditions. In the present scene rotation experiment, however, the response time was shorter than what the shortest-pass hypothesis predicts (the angle of rotation of the shortest-pass rotation was about 100 degrees). This result suggests that the underlying

process for the double axis rotation has nothing to do with the shortest-pass rotation of 3D representation. Rather, the comparison stimulus of the double axis rotation condition was processed in a similar way in the X70 condition because the sizes and shapes of the 2D images in the two conditions were similar to each other (see X70 and Y70-X70 in Figure 1).

No-Eye-Movement Trials. There were 25 trials in which the participant did not make an eye movement. Seventeen out of the 25 no-eye-movement trials were made by two participants. Thus, the no-eye-movement trials might be a result of the strategies of only a few individuals. Nevertheless, the pattern of the no-eye-movement trials was interesting. First of all, all of the non-eye movement trials were correct. Averaged over all no-eye-movement trials, the mean response time was 1191 ms. Twenty-two out of 25 no-eye-movement trials occurred in the 000, Z35 and Z70 viewpoints. In terms of the type of trial, twenty-one no-eye-movement trials were in the different-scene trials, mainly the location-3 and orientation-3 trials. Although the generality of these results are limited, the results seem to indicate that scene rotation task can be solved without eye movements if (a) the comparison stimulus was roughly 2D identical to the standard and (b) the change itself was large (i.e., all three objects were changed)

Error Trials. Out of 132 error trials, 41 were made in the same-scene condition, nine in the location-2 condition, three in the location-3 condition, 76 in the orientation-1 condition, and three in the orientation-3. To examine the difference in eye movements between the error trials and correct trials, the error trials in the orientation-1 condition were analyzed. Although the orientation-1 condition had the largest number of error

trials, the number of samples in each condition varied considerably. Thus, indices were averaged over all viewpoints and no statistical test was performed.

The largest difference was observed in the first pass time. The first-pass time in the error trials, 907 ms, was 116 ms shorter than that in the correct trials, 1023 ms. The second-pass time in the error trials, 593 ms, was also slightly shorter than that in the correct trials, 627 ms. The number of gazes in the error trials, 4.06, was slightly larger than that in the correct trials, 3.78. Also, all objects were more likely to be visited in the error trials than in the correct trials; the probability that all objects were visited was 0.74 in the error trials and 0.66 in the correct trials. The changed object was equally well visited in both error and correct trials; the probability of changed object visited was 0.93 in the error trials and 0.98 in the correct trials. The results suggested that errors were primarily caused by the eyes not spending enough time on the first-pass, rather than by a premature termination on the second pass.

Discussion

Model 3. The pattern of the data in Experiment 1 supported Model 1, which assumed the same number of gazes for the location-change and orientation-changed object. The result showed that the degrees of match/mismatch form each gaze does not accumulate faster in the location-change conditions than the orientation-change conditions. However, there was a quantitative gap between the prediction of Model 1 and the actual data; the expected number of gazes was smaller than the observation. The expected number of gazes is, 3, 1.33, 1, 1.67, and 1, and the observed value was, 4.43, 3.76, 3.63, 3.75, and 3.78 for the same-scene, location-2, location-3, orientation-1, and orientation-3 conditions respectively. The gap is remedied relatively easily. First, the

underestimation in the same-scene condition can be dealt by changing the maximum number of the 'visited' counter. In Model 1, a 'same' key press is made if all the objects were visited (i.e., the 'visited' counter shows the maximum, 3) and judged 'unchanged'. However, the maximum number may not strictly be the number of all objects. Instead, one or two more than that may be employed because the participants were careful and took an extra gaze before the key press. With this new maximum number, the expected number of gazes in the same scene condition becomes 4 to 5. Second, the underestimation in the different scene conditions can be explained if the 'different' key press is made after all objects were visited. Model 1 assumed that a 'different' key press is made immediately after an object was judged 'changed'. Instead, the participants might hold the response until all the objects were checked. In this case, the expected number of the gazes will be about the same (3) for all different-scene conditions; this matches to the pattern of the observed value. For the sake of convenience, this modified version of Model 1 is called *Model 3*.

Pre-Eye-Movement Phase. Analyses of the fixation-0 time, first-pass time and second-pass time, showed three qualitatively different patterns, suggestive of three different phases: a pre-eye-movement stage, an encoding stage, and an alignment stage. The pre-eye-movement phase is literally a period of information processing from the onset of stimulus to the start of the first saccade. The fixation-0 time, which is the main index of the phase, showed the same pattern for all stimulus types; the fixation-0 times for the 000, Z35 and Z70 viewpoints were longer than those for the X35, X70, Y35, Y70, and Y70-X70 viewpoints. This result suggests that information processing in the pre-eye-movement phase is (a) sensitive to the difference between the two groups of the

viewpoints, and is (b) common for all stimulus types. One hypothesis to explain this somewhat anomalous pattern is as follows. First, the initial information that participants extract from a comparison stimulus before eye movements is a 2D configuration of the scene. (The 2D configuration is 'crude' in the sense not being processed extensively, but needs to be precise for following processes, namely preparing the first saccade.) The 2D configuration is matched with the mental representation of the standard scene. Because the Z-axis rotation is a picture plane rotation, the 2D configuration of the comparison stimuli in the 000, Z35, and Z70 viewpoints matches well to the mental representation of the standard scene. As a result of this successful matching process, the participant may keep the 2D configuration based matching process going and may not be compelled to start moving the eyes immediately. On the other hand, when the comparison stimulus is different from the standard in 2D configuration (as it would be in the other conditions), the matching process may signal "mismatch", which may in turn signal the eyes to start moving.

However, it is rather unclear how the benefit of the 2D-configuration-based matching has an influence on the processes that follow it. For example, the eyes did not go to a changed object immediately from the initial fixation in the 000, Z35 and Z70 conditions. Averaged over the 000, Z35 and Z70 conditions, the probability that the eyes moved to a changed object directly from the fixation circle was 0.67 in the location-2 condition and 0.33 in the orientation-1 condition. Both are just slightly less than the average of the X35, X70, Y35, Y70 and Y70-X70 conditions, 0.69 in the location-2 condition and 0.40 in the orientation-1 condition. The changed object was not reached earlier in the 000, Z35 and Z70 conditions either. The number of gazes until a changed

object is reached was 1.32 in location-2 condition (averaged over the 000, Z35 and Z70 conditions), which was actually slightly more than the average of other conditions, 1.30 (averaged over the X35, X70, Y35, Y70 and Y70-X70 conditions).

There is a possibility that the underlying processes in the pre-eye-movement phase facilitated processes in the encoding phase in the Z-axis rotation condition. The first-pass time in the Z-axis rotation increased as the angle of rotation became larger. But first-pass time in the other axes was not affected by the angle of rotation. It suggests that the encoding phase finished and the alignment process surfaced earlier in the Z-axis rotation conditions than in the other rotation conditions. This might be because the 2D-configuration-based matching process gave a head start to processing in the Z-axis rotation conditions. However, how exactly the process in the pre-eye-movement phase relates to the processes in the encoding phase is not clear. (More about the role of processing in the pre-eye-movement phase is discussed in the General Discussion.)

Encoding Phase and Alignment Phase. The encoding phase is an early part of the scene processing that was not affected by the rotation, except possibly for Z-axis rotations. The encoding phase was the longest phase in the entire process; it continued for almost one second for the X- and Y-axis rotations. In contrast, the alignment phase was the late part of the scan where the effect of rotation was shown most clearly. For all axes of rotation, the second-pass time increased about 250 ms, which accounts for most of the rotation effect in the response time. It is puzzling why the rotation effect was not observed in the encoding phase even though it was the longest phase.

The main process during the first pass might be general encoding. 'General encoding' could include many processes, such as object identification, preparation for

alignment, and detection of gross differences (without attempting serious alignment). The general encoding processes are reflected in eye movement control; the more complicated the scene, the longer the eyes stay in a region to complete the general encoding. In other words, the 'eye-time' necessary for the general encoding is not a function of the angle of rotation, but a function of complexity of the scene. The general encoding process took longer in the X- and Y-axis rotation conditions than in the Z-axis rotation conditions. In other words, the piecemeal alignment started late in the former conditions than in the later. This may be the reason why the first-pass time did not increase with the angle of rotation, except for the in the Z-axis rotation conditions

To test how general the account is, the results of Just and Carpenter's cube rotation task (1985) were re-examined. The cube stimuli, black-and-white line drawings, are arguably simpler than the scene stimuli. Just and Carpenter divided the 'gaze' data of the cube rotation task into an early part and late part, but the classification and summation of gaze durations were different from those of the scene rotation task. However, relatively speaking, the "search" and "initial rotation" stages were earlier than the "confirmation" and "subsequent rotation" stages. The mean gaze duration of the late stages increased more than that of the early stages when the angle of rotation was increased (see Figure 5 in Just and Carpenter, 1985). This is consistent with the pattern observed in the present scene rotation task. At the same time, the mean gaze duration in the early stages in the cube rotation task did increase as a function of the angle of rotation. This suggests that the general encoding finished earlier in the cube rotation task than in the scene rotation task, because the cube stimuli were simpler than the scenes.

Summary. The data of Experiment 1 allow a detailed breakdown of a "mental rotation function" in the scene rotation task. During a trial, the eyes stay in an object region longer (i.e., the *gaze duration* increased) when the comparison scene was rotated more. Interestingly, gaze duration started increasing only after about 900 ms of processing. These data suggest that (a) the alignment process starts after general encoding is completed, and (b) an alignment process in the scene rotation task is piecemeal and takes place on a gaze-by-gaze basis.

As in Nakatani et al (submitted), the slope of the mental rotation function was different between conditions. Response times in the orientation-change condition increased more with increasing angle of rotation than those in the location-change condition. The difference was mainly observed in the mean gaze duration: mean gaze duration in the orientation-change condition increased more with the increasing angle of rotation than in the location-change condition. In addition, the response times in the Y (vertical)-axis rotation conditions were longer than those in the X (horizontal)- and Z (line-of-sight)-axis rotation conditions. This was chiefly because participants made a larger *number* of gazes to the objects in the Y-axis rotation conditions than the other conditions.

These results and further details of the data suggest several underlying processes and their interaction. They were summarized as Model 3. As soon as a comparison stimulus is presented, a 2D configuration of the scene is extracted before the eyes start moving. The 2D configuration appeared to be matched with the 2D mental representation of the standard scene as a preliminary process prior to the piecemeal alignment and comparison, perhaps to compute the scene rotation angle and direction. In

the next stage, the eyes are sent to one of the objects to scan the scene. At the beginning of the scan, general encoding processes dominate the eye movement control. Gradually, alignment and comparison become the main determinant, thus, what we see as a "mental rotation function" in the scene rotation task is chiefly generated during the late phase of the scan. The data also showed that the participants were cautious; they visited all objects before both 'same' and 'different' key presses, and even took an extra gaze before a 'same' key press.

In addition, Experiment 1 suggested that eye movements were guided by parafoveal vision in the location-change condition, but not in the orientation-change condition; a location-changed object was fixated earlier than an unchanged object. This phenomenon occurred during the encoding phase. (The encoding phase has about 2.5 gazes, and the location-changed object was reached around 1.4 gazes). Thus, it would be reasonable to assume that one of the general encoding processes, such as detection of gross changes, caused the phenomenon. However, details of the process are not clear. One possibility is a change in the objects relative to each other (e.g., Object 1 and Object 2 changed locations relative to each other). Another possibility is a change relevant to the local reference frame, such as the desktop (e.g. Object 1 was near by the pink-colored edge, but it is near by the white-colored edge, thus Object 1 changed its location). These issues were investigated together with others in Experiment 2.

CHAPTER 3

EXPERIMENT 2

Introduction

In Experiment 2, two possible information sources were examined which would help explain why the eyes reached a location-changed object earlier than an unchanged object. Model 3, which was modified from Model 1, was used as a guide. Object-identification studies (Pollatsek et al., 1984; Henderson et al, 1987) showed that information about the identity of an object to be fixated next are processed before the eyes moved to the object, and thus that covert attention plausibly moves to the next object before the eyes go there. Hence, it would be reasonable to assume that identity information from the to-be-fixated-next object is used to guide the eyes to a location-changed object earlier. On the other hand, non-object parafoveal information might also contribute. For example, information about the desktop is also available from the parafovea when the eyes are fixating on an object, and some desktop information could be useful for detecting location changes. For example, a mug was placed near the pink-colored edge of the desktop in the standard scene. If the mug is now next to the white-colored edge in the comparison scene, which can possibly be detected in parafoveal vision, it indicates a change in location of the mug on the desktop.

To test which source of information is more important in guiding the eyes to reach the location-changed object earlier, the desktop was removed from the scene 50% of the time in Experiment 2 (from both the standard and comparison scenes). In the *no-desk condition*, the desktop was removed from the scene (no change was made on the objects). Thus, the objects were simply placed against a black background. As a control,

scenes with the desktop were also used (the *with-desk condition*). The scenes in the with-desk condition were the same as those in Experiment 1. If object information was the main source of guidance to a changed object, the eyes should reach a location-changed object earlier than an unchanged object regardless of the presence or absence of the desktop. On the other hand, if the desktop is the main source of the benefit, the eyes shouldn't reach a changed object any earlier than an unchanged object in the no-desk condition.

The removal of the desktop also allows some details of Model 3 to be tested. The fixation-0 time data in Experiment 1 suggested that some kind of preliminary matching process specific to the 000, and Z-axis rotation conditions took place during the pre-eye-movement phase. Another process that Model 3 assumes for the pre-eye-movement phase is preparation for eye movements, such as computing object positions (or 'regions of interest', in more general terms). When the desktop is present, the objects need to be separated from the desktop. However, if there is no desktop, there is no need for the segmentation. Thus, the computation of the object position might be easier in the no-desk condition than in the with-desk condition. If this is the case, the fixation-0 time will be shorter in the no-desk condition than in the with-desk condition.

Model 3 also assumed that various general encoding processes (e.g., object identification and detection of gross change) take place in the encoding phase. Since the same objects were used in the no-desk and with-desk conditions, there should not be much of difference between the two conditions during the encoding phase. The encoding phase in the no-desk condition might end slightly earlier than that in the with-desk condition because the absolute amount of visual information needs to be encoded is less

in the no-desk condition than in the with-desk condition. Therefore, Model 3 predicts that the first-pass time, which is the index of the encoding phase, either will be the same for the both the no-desk and with-desk conditions, or that the first-pass time in the no-desk condition will be slightly shorter than that in the with-desk condition.

It is not clear how the deletion of the desktop would affect processes in the alignment phase. Model 3 assumes that a piece of a comparison stimulus is sampled gaze-by-gaze, and then aligned and compared to the mental representation of the standard scene. When the desktop is removed, the amount of visual information available for the alignment and comparison decreases. Thus, the quality of the comparison computation in the no-desktop condition may not be as good as that in the with-desk condition. Thus, a same/different judgment may be less accurate in the no-desk condition than in the with-desk conditions. Experiment 1 also showed that the participants were cautious and visited all objects before making the 'same' or 'different' key press. If the quality of the comparison is degraded, participants may make some extra gazes. Thus, the number of the gaze may be larger in the no-desk condition than in the with-desk conditions. Alternatively, the eyes might stay longer at each object in the no-desk condition than in the with-desk condition to compensate for the insufficient visual information with more careful analysis. In that case, the gaze duration in the second-pass will be longer in the no-desk condition than in the with-desk conditions[5].

[5] The second-pass gaze duration probably needs to be used because the first-pass gaze duration did not reflect alignment process in the scene rotation task in Experiment 1.

Method

Participants. Eighteen undergraduate students of the University of Massachusetts, five men and 13 women, participated to the experiment. All of the participants had normal or corrected-to-normal vision. They received either $8.00 or experimental credits in psychology courses for participating.

Stimuli and Design. Two versions of computer-generated images, scenes with the desktop and scenes without the desktop, were used in Experiment 2. The images in the with-desk condition were exactly the same as those in Experiment 1. The same objects were also used in the no-desk condition, but the square desktop was not included. Thus, the objects were placed directly against a uniform black background. The five stimulus types (i.e., same-scene, location-2, location-3, orientation-1, and orientation-3) and the eight viewpoints (i.e., 000, X35, X70, Y35, Y70, Z35, Z70 and Y70-X70) were exactly the same as those in Experiment 1.

Procedure. The procedure for Experiment 2 was the same as that for Experiment 1 except that the subjects participated a two-day session to complete both the with-desk condition and the no-desk condition. The experimental session was split into two days to minimize the physical strain caused by the task and the head-mounted eye tracking system. On the first day, the participants were given 32 practice trials with feedback, then finished four blocks of 64 trials without feedback. As in Experiment 1, there were three office objects in a standard scene, and one standard scene was used for each block. Thus, there were four standard scenes. On half of the four blocks, the desktop was removed (no-desk condition), but on the other half of the blocks, the desktop was present (with-desk condition). Therefore, the participants completed a half of the no-desk trials

and a half of the with-desk trials on the first day. On the second day, the rest of the no-desk condition and the with-desk condition were completed. The order of the no-desk and with-desk conditions was counterbalanced. For example, if a participant was given Scene 1 and Scene 2 in the no-desk condition and Scene 3 and Scene 4 in the with-desk condition on the first day, he or she had Scene 1 and Scene 2 in the with-desk condition and Scene 3 and Scene 4 in the no-desk condition on the second day. Moreover, half the subjects started with the no-desk condition and then moved to the with-desk condition on Day 1, with the order of the no-desk and with-desk conditions reversed on Day 2. The order was reversed for the other half of the subjects. The mean interval between the two sessions was 5.22 days.

The eye movements during the trials were recorded by the EyeLink system. The procedure of the recording was the same as that in Experiment 1.

Results

Excluded data. One of the participants reported that she tried deliberately not to move her eyes, and her data were excluded from analyses. Thus, the following analyses were conducted on the data from the other 17 participants. As in Experiment 1, error trials were excluded from computations of the eye movement indices. There were 155 error trials in the with-desk condition (3.56 %) and 273 in the no-desk condition (6.30 %). There were also 65 no-eye-movement trials (1.49 %) in the with-desk condition, and 34 (0.78 %) in the no-desk condition. The no-eye-movement trials were also excluded from the main analysis, but are discussed separately. Moreover, the fixations made outside of the display were excluded from the computation of eye movement indices.

The number of the out-of-screen fixation was 320 (1.24 %) in the with-desk condition and 306 (1.10 %) in the no-desk condition.

In the next section, response times, error rates, and various eye movement indices were examined for both the no-desk and with-desk conditions. The main eye movement indices were the same as those in Experiment 1: fixation-0 time (the latency to the first saccade), first-pass time (the sum of fixation durations before the first regression), second-pass time (the sum of fixation durations after the first regression), gaze duration (the sum of fixation durations to an object per visit) and the number of gazes. The data in the double axis rotation condition (Y70-X70) are discussed after the single axis rotation conditions (000, X35, X70, Y35, Y70, Z35, and Z70).

Response times and errors. The response times in the no-desk condition were slightly longer than those in the with-desk condition, and the difference was mainly due to the Y-axis rotation condition (see Table 5). The main effect of the presence/absence of the desktop was not significant, $F(1, 16) = 1.98$, $p > .1$[6], but the interaction of the presence/absence of the desktop and the axis of rotation was, $F(2, 32) = 4.61$, $p. < .002$. The difference between the no-desk and the with-desk condition was about 250 ms in the Y axis rotation condition, $F(1, 16) = 8.07$, $p < .02$, but small (less than 50 ms) for the X- and Z- axis rotations, $Fs < 1$. The error data showed that the participants was less accurate in the no-desk condition than in the with-desk condition, $F(1, 16) = 21.63$, $p < .001$, for the main effect of the presence/absence of the desktop. As with the response

[6] The experimental design was not completely factorial. Because there is only one 0-degree condition for three axes of rotations. Thus, the 0-degree condition was excluded from the 2x3x2 ANOVA (presence/absence of the desktop by axis of rotation by 35- or 70-degree of rotation) used in this paragraph.

times, the difference was significant for the Y-axis rotation condition, $F(1, 16) = 10.24$, $p < .001$, but not for either the X- and Z-axis rotation conditions, $F(1, 16) = 2.62$, $p > .1$, $F < 1$, respectively. Thus, both response time and errors showed that processing the Y-axis rotation became difficult when the desktop was removed, whereas the X- and Z-axis rotation conditions were not affected by the presence/absence of the desktop very much.

The intercepts and slopes of the 'mental rotation function' were about the same for the no-desk and with-desk conditions. The no-rotation (0-degree) condition of the no-disk condition was slightly slower and less accurate than that in the with-desk condition, but the differences were not significant, ts < 1 for the response times and error rates. Thus, the intercepts of the mental rotation function of the two conditions were about the same. In both the no-desk and with-desk conditions, the response times increased as the angle of rotation increased from zero to 70 degrees. The slopes in both conditions were significantly greater than zero for all axes of rotation, $ts(16) > 2.46$, $ps < .025$. The slopes in the no-desk conditions were greater than those in the with-desk conditions, but neither the main effect of the presence/absence of the desktop nor the interaction between the presence/absence and the axes were significant, $F(1, 16) = 1.26$, $p > .1$, $F < 1$[7], respectively. In the error rate data, however, the slopes in the no-desk condition were larger than those in the with-desk, $F(1, 16) = 5.62$, $p < .05$. The difference between the no-desk and with-desk condition was significant for the Y-axis rotation condition, $t(16) = 2.33$, $p < .05$, but not for the X- and Z- axis rotation conditions, $t(16) = 1.40$, $p > .1$, $t < 1$, respectively. In short, the presence/absence of the desktop did not greatly affect the

[7] A 2x3 ANOVA (presence/absence by three axes) was applied to the 0-70 slopes.

'mental rotation function' in the response times. On the other hand, the slope in the error rates increased significantly when the desktop was removed, but only in the Y-axis rotation condition. These results also suggest that the effect of the removal of the desktop was concentrated in the Y-axis rotation condition.

The response times and error rates in the same-scene and four different scenes were listed in Table 6. The effect of the removal of the desktop was observed more clearly in the error rates than in the response times. In the same-scene condition, the error rates were larger in the no-desk conditions than in the with-desk conditions, but the response times were about the same for both the no-desk and with-desk conditions: averaged over the seven viewpoints, $t(16) = 4.59$, $p < .001$, $t < 1$, for error rates and response times, respectively. Error rates increased particularly for Y-axis rotations; from 2.03 % to 19.31 % in the Y-axis rotation conditions (+17.28% averaged over 35- and 70-degree conditions), $F(1, 16) = 25.34$, $p < .001$[8]. In contrast, the increments were only +1.65 % in the X-axis rotation conditions, and +0.73% in the Z-axis rotation conditions ($Fs < 1$).

The effect of the removal of the desktop was larger in the location-change conditions than in the orientation-change conditions, but the difference was clearer in the error rates than in the response times. For the location-change conditions, the mean response time in the no-desk condition was 166 ms more than in the with-desk condition (averaged over the seven viewpoints of the location-2 and location-3 conditions) but, $t(16) = 1.35$, $p > .1$. However, the 1.47% difference in the error rates was significant,

[8] The main effect of presence/absence of the desktop of a 2x2 ANOVA (presence/absence by 35-70 degrees) that was applied to each axis of rotation.

$t(16) = 2.14$, $p < .05$. In the orientation-change conditions, the average response time was 109 ms longer in the no-desk condition than in the with-desk condition (averaged over the seven viewpoints in the orientation-1 conditions and orientation-3 conditions), $t(16) = 1.40$, $p > .1$, but the error rate in the no-desk condition, was actually .84 % less than that in the with-desk condition, but $t < 1$. In short, the location-change condition seemed to be affected negatively by the removal of the desktop more than the orientation-change condition was; when the desktop was removed, the accuracy decreased in the location-change trials, but not in the orientation-change conditions.

The Number of Gazes to Reach a Location-Changed Object. The object to be fixated next and the desktop were considered two possible information sources helping to send the eyes earlier to a location-changed object than to an unchanged object. If the object to be fixated next is the main source, the number of gazes until the eyes reach the changed object should be smaller than that to an unchanged object regardless of the presence/absence of the desktop. On the other hand, if the desktop was the main source, the number of gazes to the changed object and that to the unchanged object should be the same in the no-desk condition.

As in Experiment 1, the number of gazes to a changed object (e.g., mug) in the location-2 and orientation-1 conditions was compared to the number of gazes to the object (mug) in the same-scene condition. In the with-desk condition, the number of gazes to a location-changed object (1.38 gazes, averaged over seven viewpoints) was smaller than that to the same object in the same-scene condition (1.61 gazes), $t(16) = 4.03$, $p < .002$. However, in the no-desk condition, the number of gazes to a location-changed object (1.40 gazes) was about the same as that of the same-scene condition (1.46

gazes), $t(16) = 1.15$, $p > .1$. A 2x2 ANOVA (presence/absence of the desktop by location-2 vs. same-scene) showed that the interaction was also significant, $F(1, 16) = 5.68$, $p < .05$. Thus, the results suggested that the parafoveal information that guided the eyes to the location-changed object earlier than to an unchanged object was chiefly provided by the desktop (or the relationship of the object with the desktop).

Effect of The Removal of the Desktop in the Pre-Eye-Movement, Encoding and Alignment Phases. Model 3 predicted that the fixation-0 time of the no-desk condition should be shorter in the no-desk condition than in the with-desk condition, because the objects do not need to be 'segmented' from the background in the no-desk condition. In fact, when the fixation-0 times of the seven viewpoints were averaged, the average fixation-0 time in the no-desk condition (240 ms) was 16 ms less than that in the with-desk condition (256 ms), $t(16) = 2.55$, $p < .025$. The results suggest that the processes necessary to start moving the eyes were completed earlier in the no-desk condition than in the with-desk condition.

The qualitative pattern of the fixation-0 times, however, was the same in the no-desk and with-desk conditions (see Table 7). The two-group pattern was seen in both conditions; the fixation-0 times in the 000, Z35 and Z70 viewpoints were about 50 ms longer than those in the X35, X70, Y35 and Y70 viewpoints. There was no significant difference among the 000, Z35 and Z70 conditions, $F(1, 16) = 1.69$, $p > .1$, $F(1, 16) = 1.94$, $p > .1$ for the no-desk and with-desk conditions respectively. The average fixation-0 time of the 000, Z35 and Z70 conditions was 216 ms in the no-desk condition, and 237 ms in the with-desk condition. The average fixation-0 time in the 000, Z35, and Z70 conditions was longer than the fixation-0 time of the X35, X70, Y35 and Y70 conditions

for both the no-desk and with-desk conditions, $ts(16) > 2.66$, $ps < .05$, $ts(16) > 3.07$, $ps < .05$, respectively (probabilities were adjusted by using the Bonferroni method). The results indicated that the qualitative aspect of the underlying processes during the pre-eye-movement phase (i.e., preliminary matching specific to 000, Z35 and Z70 condition) did not change between the no-desk and with-desk conditions, but the processes finished earlier in the no-desk condition than in the with-desk condition. It appears that segmentation of the objects from the desktop was not necessary.

The first-pass times showed about the same pattern in the no-desk and with-desk conditions, which is in agreement with the prediction of Model 3. When the first-pass times were averaged all viewpoints except for the double axis condition, the average first-pass time in the no-desk condition (837 ms) was about 30 ms faster than that in the with-desk condition (870 ms), but the difference was not significant, $t(16) = 1.50$, $p > .1$. When the angle of rotation increased, the first-pass times increased only in the Z-axis rotation condition; the slope for the Z-axis conditions was significantly greater than zero, but those in the X- and Y-axis rotation conditions were not, $F < 1$, $F < 1$, $F(1, 16) = 5.95$, $p < .05$, for the X-, Y- and Z- axis respectively[9]. The slope in the Z-axis rotation was about the same in the no-desk and with-desk conditions, $F < 1$ for the interaction between the presence/absence and 0-70 degrees. In Experiment 1, the first-pass time also increased only in the Z-axis rotation condition when the angle of rotation increased. This pattern suggests that the alignment phase started earlier in the Z-axis condition than in the X- and Y-axis conditions as a result of the early completion of the encoding phase.

[9] The 2x2 ANOVA (presence/absence by 0-70 degrees) was applied to each axis of rotation.

In sum, the first-pass time data showed that the encoding phase was not affected by the removal of the desktop very much.

In contrast, the second-pass times were significantly affected by the presence/absence of the desktop, especially in the Y-axis rotation conditions. When the second-pass times in the seven viewpoints were averaged, the average second-pass time in the no-desk condition (793 ms) was 120 ms longer than that in the with-desk condition (673 ms), $t(16) = 2.45$, $p < .05$. In both the no-desk and with-desk conditions, the second-pass times increased for all axes of rotation between zero and 70 degrees, $F(1, 16) = 9.64$, $p < .01$, $F(1, 16) = 28.83$, $p < .001$, $F(1, 16) = 18.25$, $p < .002$[10], for the X-, Y- and Z-axis rotations respectively. The increase was the largest in the Y-axis rotation of the no-desk conditions – twice as much as in the X- or Y- axis conditions, $t(16) = 2.01$, $t < .1$ for Y- vs. X-axis, $t(16) = 2.96$, $p < .01$ for Y- vs. Z-axis.

The second-pass time increased in the no-desk condition chiefly because the number of gazes increased. Averaged over the seven viewpoints, the number of gazes in the second-pass was 2.09 in the no-desk condition and 1.75 in the with-desk condition, $t(16) = 3.09$, $p < .01$. When the number of gazes was examined for each axis of rotation in the no-desk condition, the number of gazes increased the most in the Y-axis (+0.62 gazes), the second most in the X-axis (+0.30), and the least in the Z-axis (+0.18) rotation conditions. A 2x3x2 ANOVA (no-desk vs. with-desk, axes of rotation and 35- vs. 70-degree) showed that the 2-way interaction between the presence/absence of the desktop

[10] The 2x2 ANOVA (presence/absence by 0-70 degrees) was applied to each axis of rotation.

and the axes of rotation was marginally significant, $F(2, 32) = 3.04$, $p = .07$[11]. These results suggest that more gazes were made when the desktop was removed to compensate for the loss of the desktop information. Also, the loss of the desktop tended to increase the number of gazes the most in the Y-axis rotation condition.

On the other hand, the mean second-pass gaze duration was about the same in the no-desk and with-desk conditions, 330 ms and 335 ms, respectively (averaged over the seven viewpoints, $t < 1$, $p > .1$). However, the pattern of the gaze duration was somewhat different between the two conditions. As in Experiment 1, the gaze duration more or less increased for the greater angle of rotation in the with-desk condition; the differences between the 0- and 70-degree conditions were 22 ms, 30 ms, and 48 ms for the X-, Y- and Z-axis respectively. In contrast, in the no-desk condition, the differences were –11 ms, 3 ms, and 17 ms for the X-, Y- and Z-axis, respectively. However, in a 2x3 ANOVA (no-desk vs. with-desk, and axes of rotation) on the differences, both the main effect of the presence/absence of the desktop and interaction with axis of rotation were not significant. Given the non-significant result, it is difficult to conclude what the data exactly mean, but the gaze durations suggested that the eyes tend to stay longer in the same object region for the greater angle of rotation when the desktop was present.

Double Rotation Condition. The response times in the double axis rotation condition were longer in the no-desk condition (2320 ms) than in the with-desk condition (2080 ms), $t(16) = 2.43$, $p < .05$. The error rate was also higher in the no-desk condition (5.70 %) than in the with-desk condition (4.78 %), but the difference was not significant,

[11] The main effect of the presence/absence of the desktop was also significant, $F(1, 16) = 10.27$, $p < .01$.

$t < 1$. The eye movement indices in the double axis rotation condition showed a similar pattern to that in the single axis rotation conditions. The fixation-0 time was slightly shorter in the no-desk condition (223 ms) than in the with-desk condition (230 ms), but $t(16) = 1.03$, $p > .1$. The first-pass time was also shorter in the no-desk condition (865 ms) than in the with-desk condition (877 ms), but $t < 1$. In contrast, the second-pass time was almost 200 ms longer in the no-desk condition (822 ms) than in the with-desk condition (635 ms), $t(16) = 3.24$, $p < .01$. In both the no-desk and with-desk conditions, the overall pattern of results in the double-axis condition was more similar to the X70 condition than to the Y70 condition in Experiment 1 (see Table 5). Thus, the removal of the desktop did not affect the relationship among the double-axis and X70 and Z70 conditions.

No-Eye-Movement Trials. In Experiment 2, there were also a small number of trials in which the participants did not move their eyes (i.e., no-eye-movement trials). As in Experiment 1, there were differences between individuals; some participants had many no-eye-movement trials, others didn't. Thus, no statistical test was performed on the data, but the pattern of data is of interest. The no-eye-movement trials occurred about twice as often in the with-desk condition (65 trials) as in the no-desk condition (34 trials). All of the no-eye-movement trials were correctly responded to except for one trial in the no-desk condition. In Experiment 1, the majority of no-eye-movement trials occurred in the 000, Z35 and Z70 conditions. However this was not the case in Experiment 2. In the with-desk condition, only about a half of the no-eye-movement trials (32 trials) occurred in the 000, Z35 and Z70 conditions, and in the no-desk condition, less than a half of the no-eye-movement trials (14 trials) took place in these conditions. Thus, the process

behind the no-eye-movement trials might not be a simple 2D template matching. In terms of type of change, the no-eye-movement trials occurred in trials with a 'large' change (i.e., location-3 and orientation-3 conditions) more frequently than in trials with a 'small' change (i.e., location-2 and orientation-1 conditions). In the with-desk condition, there were 26 no-eye-movement trials in the large change conditions, 15 in the small change conditions, and 24 in the same-scene condition. In the no-desk condition, there were 15 in the large change conditions, 6 in the small change conditions and 12 in the same-scene condition. Thus, the data suggest that a change in the scene could be processed without an eye movement when the change was large.

In addition to the frequency of the no-eye-movement trials, the response times in the no-eye-movement trials were checked in each viewpoint. To have the maximum number of no-eye-movement trials, those in the with desk conditions and in Experiment 1 were analyzed together. Thus, each of the viewpoints had more than 10 no-eye-movement trials to be averaged, except for the Y35m Y70 and Y70-X70 conditions (7, 5, and 4 trials respectively). The response times increased for a greater angle of rotation in all axes of rotation. Averaged over all available trials, the response times were, 967 ms (000), 992 ms (X35) and 1031 ms (X70), 938 ms (Y35), 1613 ms[12] (Y70), 993 ms (Z35), 1104 ms (Z70), and 1086 ms (Y70-X70). Thus, the results suggest an alignment (and comparison) can happen without eye movements. In the no-eye-movement conditions, the alignment might not be applied to any of the objects specifically because the eyes are not aimed to any of the objects. It is rather plausible to that the alignment was applied to

[12] A response time of the no-eye-movement trials in the Y70 condition was anomalously long (4431 ms). Thus, the average of the Y70 condition might be overestimated.

a 2D configuration of an entire scene – it makes sense because the 2D configuration was extracted during the pre-eye-movement. The alignment of an entire scene may be less accurate than a piecemeal alignment, but it may be sufficient to detect the large changes (i.e. changes in the location-3 or orientation-3 conditions).

Discussion

Role of the Desktop in the Scene Rotation Task. Experiment 2 showed that the desktop information was used (a) to guide eyes earlier to the location-changed object than an unchanged object and (b) to detect a location-change (the error rates in the location-change conditions increased when the desktop was removed). Thus, the desktop was more important in the location-change condition than in the orientation-change condition. This makes sense because the location changes can be detected by the proximity of an object to salient features of the desktop, such as the edges.

The effect of the removal of the desktop was seen chiefly in the second-pass time, especially in the Y-axis rotation. The second-pass time increased mainly because the participants made more gazes in the no-desk condition than in the with-desk condition. Despite the increased number of gazes, the error rates were also the highest in the Y-axis rotation condition. These results make sense intuitively. Without the desktop, the object locations do not have a rigid frame of reference within a scene. Thus, in the Y-axis rotation condition, alignment of the objects may be difficult because the left-right relationship among the objects is changed. The Y-axis effect can be explained more specifically with Model 3. A process directly responsible for the Y-axis specific effect is the correspondence process between a visual input from a gaze and the mental representation of the standard scene. Model 3 (as well as Model 1) assumes that a piece

of the comparison scene captured by a gaze is placed on a corresponding part of the mental representation of the standard scene *prior* to an alignment. For the correspondence process, whatever available features sampled in the gaze are used. In the no-desk condition, the amount of available information that can be used by the correspondence process is simply less than that in the with-desk condition. For example, in the with-desk condition, features of the desktop (e.g., pink-colored edge) are available in addition to object information. This additional desktop information is useful in establishing a correct correspondence. As a result, all of the following processes, the alignment, the comparison, and the same/different judgment, will suffer.

When the desktop is absent, a correspondence must be established based on non-desktop information. For example, a 2D configuration extracted in the pre-eye-movement phase might be used for the correspondence process. The 2D configuration, a triangle made by the three objects, may be enough to establish the correct correspondence in the X- and Z-axis rotation conditions, but not in the Y-axis rotation condition. In the Z-axis rotation conditions, the triangular configuration of the objects of the comparison stimuli was more or less the same as that of the standard scene. In the X-axis rotation condition, the configuration is still a valid cue for the initial correspondence although it needs to be stretched in the vertical axis. However, in the Y-axis rotation condition, the configuration is skewed more than a simple elongation of the triangle. Thus, it is difficult to establish the correct correspondence based solely on the triangle configuration. As a result of the poor initial correspondence, the subsequent processes are likely to suffer more in the Y-axis rotation than in the X- and Z-axis rotation conditions. The alignment phase becomes longer because the correspondence and

alignment need to be re-done, yet, the error rate is higher in the in the no-desk condition than in the with-desk condition because of the misalignment that has already happened.

In short, the desktop appears to have two major roles. First, it provides additional information for the alignment and comparison process. Second, it is also used to help guide eyes to a location-changed object.

CHAPTER 4

GENERAL DISCUSSION

Underlying Processes in the Scene Rotation Task

When a scene is rotated, we tend to take more time to judge if the objects on a desktop are in the same location or are facing in the same direction as before. These times increase with the angle of rotation. Based on our eye movement experiments, this 'mental rotation' effect in a scene rotation task could be seen as consisting of different component parts. The first of these components is the time spent looking at each object on a "visit" to that object -- the mean gaze duration on individual objects of the scene was longer for greater angles of rotation. Second, a larger number of gazes were made when the rotation angle increased. The rotation effect in gaze duration and number of gazes are predominantly observed late during the scanning process. Experiments 1 and 2 indicated that the rotation effect in the second-pass time explains about 75 % of the entire rotation effect in response time.

Nakatani et al. (submitted) reported that the 'slope' of the 'mental rotation' function depends on the type of change that occurs in the scene structure – the slope was steeper in the orientation-change condition than in the location-change condition. The present study shows that the difference in the slope is primarily manifested in the gaze durations -- the gaze duration in the orientation-change condition increased more than in the location-change conditions for the greater angle of rotation. Thus, as most of the interesting response-time patterns in the scene rotation task are already reflected in the duration of individual gazes, it would be reasonable to conclude that the underlying processes for the scene rotation task are closely related to gaze durations.

Nakatani et al. observed that Y-axis rotations had the longest response times and highest error rates among the three single axis rotation conditions. The causes of differences among the axes of rotation appear to be rather complex and the eye movement data suggest that interactions of several underlying processes might cause the difference. In the following paragraphs, the possible underlying processes for the scene rotation are discussed. Next, an interaction of these processes that might cause the difference among the axes of rotation is described.

When a comparison stimulus is presented, two processes seem to start immediately. One is preparation for an eye movement, such as listing 'areas of interest' (e.g., the objects), and the other is a matching between a holistic 2D configuration of a comparison scene and the mental representation of the standard scene. The first process appears to be based on a 2D image of the scene[13]. The process computes positions of 'areas of interest', such as the objects (or parts of the objects) and salient features of the desktop (e.g., colored edges and corners). These 'areas of interest' are candidates where the eyes can be directed in a later stage of processing. The process may be relatively simple and data-driven. Fixation-0 time, the latency of the first saccade, decreased when the desktop was removed. This result suggests that the preparation for eye movements finishes earlier when the absolute amount of information is decreased. Manann, Ruddoch and Wooding (1995 and 1997) suggested that low spatial-frequency information is important for guiding eye movements. Manann et al. showed low-pass filtered, high-

[13] The scene stimuli are 2D. Any depth information needs to be recovered from the 2D image. Since the process discussed here is assumed at the very beginning of the processes. Thus, it might be sensible to assume that the process is based on the 2D image of the scene.

pass filtered, and unfiltered images to participants and found fixation positions in the first 1.5 sec were similar between the low-pass-filtered and unfiltered-image conditions. Thus, sub-processes for planning the location of eye movements may be rather simple – probably something like a low-pass filtering, plus computation of eye-movement vectors (Wurtz and Munoz, 1994).

The second process seems to be a matching between a holistic 2D configuration of a comparison scene and the mental representation of the standard scene. Holistic alignment and comparison are somewhat unexpected in this context, because the overall pattern of data supports piecemeal alignment. The data of the no-eye-movement trials suggested this possibility – participants were occasionally able to detect location- or orientation-changes correctly without moving the eyes if the change was 'large' (i.e., location-3 and orientation-3 changes). More importantly, the response times in the no-eye-movement trials increased as the angle of rotation increased. On the other hand, the fixation-0 times in the 000, Z35 and Z70 conditions were longer than that of the other conditions -- this suggests that some processes are keeping the eyes from moving in the 000, Z35 and Z70 conditions. It might be the case that the holistic 2D alignment and comparison were occasionally successful in the 000 and Z-axis rotation conditions because the comparison scene in these conditions shared all the 2D features of the standard scene). Thus the success of a 2D holistic alignment and matching might delay the onset of eye movements.

Interestingly, these two processes, the preparation for an eye movement and holistic alignment and matching, seem to be independent. More precisely, the preparation for the eye movement is a part of the piecemeal alignment process, and it

does not seem to be influenced by the holistic alignment and matching process. For example, the transition probability from the fixation circle to a changed object did not differ between the 000 & Z-axis rotation conditions and the other ones. If a preliminary holistic alignment process had influenced the eye movement preparation process, the eyes would have moved to a changed object more directly from the fixation circle in the 000 & Z-axis rotation conditions. Likewise, the eyes did not reach a changed object any earlier in the 000 and Z-axis conditions than in the other conditions. These data indicate that a result of holistic alignment and comparison does not affect the piecemeal alignment system that involves eye movements. In short, both holistic and piecemeal alignment and comparison processes are evoked when a comparison scene was presented. Participants, however, would need to use the piecemeal alignment process more than 98 % of the times (no-eye-movement trials occurred less than 2 % of the entire trials). Moreover, these two processes are more or less independent in the pre-eye-movement phase. (The holistic alignment and matching processes may have shortened general encoding in the Z-axis rotation conditions.)

In the piecemeal alignment process, the eyes start visiting the objects in turn as soon as the positions of the objects are extracted. In the first 2-3 gazes, it cannot be concluded whether attempted alignment and comparison is occurring – the durations of the first-pass gazes were long (more than 300 ms), but were not affected by the angle of rotation (except for the Z-axis rotation conditions). This is probably because general encoding processes, such as object identification, are dominant in eye movement control at the beginning of the scan (van Diepen, Wampers and d'Ydewalle, 1998). Later in the scan, however, the mean gaze duration and the number of gazes are a sensitive index for

the alignment and comparison processes. These indices suggest the following processes: once the eyes have moved to one of the objects, a piece of visual information is sampled. The sample is probably put in correspondence with the mental representation of the standard scene, and then is aligned and compared with it. The correspondence process is based on any information available, such as parts of the object, the desktop features, and the 2D configuration of the scene (e.g., the triangular configuration of the objects). When a correspondence is erroneous, all subsequent processes suffer. For instance, a correspondence may be more likely to be erroneous in the Y-axis rotation conditions than in the Z-axis rotation conditions, because the 2D configuration of the scene changed more for the Y-axis rotation than for the Z-axis rotation. The correspondence also is more likely to be erroneous when the desktop is removed. This is because the amount of information available for the correspondence process is decreased. As a result of incorrect correspondence, more errors were made in the Y-axis rotation and no-desk conditions. Moreover, when a correspondence is incorrect, a same/different judgment based on local alignment and comparison may yield inconsistent information from one gaze to another (e.g., the result of comparison in Gaze 3 was 'same', but that in Gaze 4 was 'different'). The inconsistency between gazes might be the cause of more gazes being needed in the Y-axis and no-desk conditions.

To summarize, there is no single process solely responsible for the differences among the axes of rotation. Rather, they may be understood from an interaction of several underlying processes in the scene rotation task.

Relation to Models of Object Recognition across Viewpoint Changes

Various models have been proposed to explain object recognition across viewpoint changes (Shepard and Metzler, 1971; Tarr and Pinker, 1989; Hummel and Biederman, 1992; Gerhardstein and Biederman, 1993; Poggio and Edelman, 1990; Perrett, Oram, and Ashbridge, 1998; Riesenhuber and Poggio, 1999). In this study, viewpoint-dependent models based on 2D representation were assumed to apply to scene rotation because: (a) recognition of the scenes is viewpoint-dependent (Nakatani et al.); and (b) some viewpoint-dependent models, such as the Gaussian radial basis function (RBF) network model, are supported by neurophysiological studies (Logothetis, Pauls, Bülthoff, and Poggio, 1994; Logothetis, Pauls and Poggio, 1995). The performance of these models, however, is not robust when multiple objects are involved (Poggio and Edelman 1990). Moreover, these models have no mechanism whatsoever to take eye movements into account – this restricts these relevance of this models with respect to the real-time interaction of the processes during the scene rotation task.

Riesenhuber and Poggio (1999) modified the Gaussian RBF framework in order to accommodate multiple-object conditions. Part of their solution was to introduce both local and global representations in their model – in their *hierarchical RBF* (hRBF) model, a 2D image is processed by multiple-layers of filters that have different sizes and orientation tuning of receptive fields - the filters were modeled after response functions of V1 to V4 neurons. With this multiple-layer filtering system, the model is able to have both representation of local features and entire objects. To prevent an incorrect clustering of local features (e.g., a feature belonging to Object 1 is assigned to Object 2), non-linear weights were used to sum up outputs from a lower layer, as well as

conventional weighted sums. After training the model, the network was able to discriminate an object from the background.

Apparently, the main emphasis of the hRBF network model is on static memory representations. These have been optimized for multiple purposes. Depending on conditions, they can now be used for viewpoint dependent or independent processing. The model, however, does not address perception as a dynamic interaction between visual inputs and memory representations. As a consequence, the hRBF network model still falls short as an approach that could explain scene rotation effects, insofar as the effect involves the integration of information from multiple gazes.

The hRBF network may be considered as a possible memory/learning architecture in conditions where sets of examples, such as various views of paper-clip objects, are presented repeatedly with a feedback signal from a supervisor. In this respect they are essentially a standard connectionist architecture (Rumelhart, Hinton, and McClelland, 1986), as was acknowledged (Poggio and Girosi, 1990). These models are far removed from an understanding of perceptual dynamics.

Similar restrictions apply, to some degree, to the population-coding approach proposed by Perrett, Oram, and Ashbridge (1998). Perrett et al. identified groups (i.e., populations of cells) in macaque temporal cortex that selectively responded to different views of a human face. Each population responded to a particular view of a face. The activation of a population gradually decreased when the face was rotated away from its preferred view. Activity in a population of cells, therefore, accumulates more slowly, the more the orientation of the face differs from the preferred orientation. Perrett et al. argued that the rotation effect in response times is not because of an alignment process,

but because of the speed of accumulation of activity in a population (reflecting degree of mismatch with a template).

These observations yield a plausible hypothesis for gaze control: the eyes start moving as soon as a certain amount of activity is accumulated from a population of temporal neurons. However, it should also be noted that this hypothesis explains just a part of the mechanism for the rotation effect in response time. It explains effects on gaze duration, but provides no mechanism for controlling gazes and fails to explain, for instance, why the number of gazes depends on the axes and angle of rotation.

There is one more reason for caution in applying the population-coding approach to the scene rotation task. Perrett et al. claimed that composite objects, such as the human figure with head and body, are represented by the same population-coding scheme as the scheme for faces. They propose that the rotation of a composite object can be dealt with in the same manner as a simpler object like a face. It is not clear, however, to what extent the scheme can be extended to the multi-object situation, such as the scene rotation task. In the case of face stimuli, the cell population in temporal cortex showed a sizable difference in the rate of activation accumulation – the population activity for the 0-degree face accumulated much faster than for the 45-degree rotated face. Careful examination of their data shows, however, that when composite patterns were used, the rate of accumulation was about the same for the 0- and 45-degree rotated stimuli (Wachsmuth, Oram and Perrett, 1994). In other words, population activity seems to lose discrimination power quickly when stimulus complexity increases. Thus, as Riesenhuber and Poggio (1999) proposed, some form of memory elaboration would be necessary in their framework to keep a reasonable descriminability among the stored views.

Conclusion

It is clear what is missing from the existing models for the object recognition over viewpoint changes: the dynamic interaction between memory representation and perception is nowhere taken into account. Eye movements are a very useful vehicle to investigate the interaction. In principle, there is no fundamental difference between the perception of a rotated object and a rotated scene – both objects and scenes have local and global features. However, rules applied to the features are often more strict in objects than in scenes (e.g., a mug can be inside a cupboard, but a handle of the mug cannot be inside the mug). Thus, further investigation of the relationship between eye movements and the structure of memory (e.g. memory storage tagged and accessed by eye movement vectors) is necessary for a more complete understanding of the perception of rotated scenes.

Table 1. Response times and error rates in Experiment 1.

All stimulus types were averaged over. Paired t-tests were performed between the 0- and 70-degree conditions.

Axis of rotation	**Degree of rotation**				**Slope in ms/deg and percent error per degree**
	0	35	70	100	
X	**RT and Error** 1937 ms 1.79%	2017 0.89	2122 2.83	--	2.64** 0.02
Y		2128 3.28	2281 3.72	--	4.90*** 0.03**
Z		2016 1.94	2293 1.94	--	4.91*** 0.00
Double		--	--	2093 3.28	1.56^ 0.02^

* $p < .1$, ** $p < .05$, *** $p < .01$, ^ no test

Table 2. Response times and error rates in Experiment 1, listed by each stimulus type.

The values are listed in the same-scene and different-scene conditions. The values in the different scene condition were further broken down to the location-2, location-3, orientation-1 and orientation-3 conditions. Paired t-tests were performed between the 0- and 70-degree conditions.

Axis of rotation (RT and error rate)	**Degree of rotation**				**Slope in ms/deg and percent error per degree**
	0	35	70	100	
Same					
X	1913ms 0.89%	2180 0.89	2219 1.79	-- --	4.37*** 0.01
Y		2325 3.27	2285 2.08	-- --	5.31*** 0.02
Z		1998 1.49	2188 0.60	-- --	3.92*** 0.00
Double		-- --	-- --	2137 1.19	2.27^ 0.00^
Different					
X	1961ms 2.68%	1855 0.89	2025 3.87	-- --	0.90 0.02
Y		1931 3.27	2276 5.36	-- --	4.49*** 0.04**
Z		2034 2.38	2375 3.27	-- --	5.91*** 0.01
Double		-- --	-- --	2050 5.36	0.88^ 0.03^
Location-2					
X	2181ms 0%	1757 1.19	1779 2.38	-- --	-5.74** 0.03^
Y		1909 2.38	2001 2.38	-- --	-2.57 0.03^
Z		2371 0	2218 1.19	-- --	0.53 0.02^
Double		-- --	-- --	1821 1.19	-3.60 0.01^

Continued Next Page

Table 2. Continued

Location-3					
X	1856ms	2295	1916	--	0.86
	0%	1.19	0	--	0.00
Y		1793	1846	--	-0.14
		1.19	0	--	0.00
Z		2036	2233	--	5.39***
		0	0	--	0.00
Double		--	--	1880	0.24^
		--	--	1.19	0.01^
Orientation-1					
X	2044ms	1754	2570	--	7.51**
	10.71%	1.19	13.10	--	0.03
Y		1747	2291	--	3.53
		9.52	16.67	--	0.09
Z		1939	2519	--	6.79**
		8.33	11.90	--	0.02
Double		--	--	2852	8.09^
		--	--	19.00	0.08^
Orientation-3					
X	1764ms	1614	1833	--	0.98
	0%	0	0	--	0.00
Y		2274	2964	--	17.15***
		0	2.38	--	0.03^
Z		1791	2529	--	10.93***
		1.19	0	--	0.00
Double		--	--	1644	-1.20^
		--	--	0	0.00^

* $p < .1$, ** $p < .05$, *** $p < .01$, ^ no-test

Table 3. Fixation-0 times, first-pass times, and second-pass times in Experiment 1

All stimulus types were averaged over. Paired t-tests were performed between the 0- and 70-degree conditions.

Axis of rotation	Degree of rotation				Slope in ms/deg
	0	35	70	100	
X	**Fixation-0 first-pass and second-pass** 276 ms 919 ms 517 ms	221 941 540	214 926 598	-- -- --	-- 0.09 1.17
Y		227 890 686	203 897 828	-- -- --	-- -0.31 4.46***
Z		280 943 533	273 987 736	-- -- --	-- 0.97** 3.14***
Double		-- -- --	-- -- --	204 975 558	-- 0.56^ 0.41^

* $p < .1$, ** $p < .05$, *** $p < .01$, ^ no-test

Table 4. Fixation-0 times, first-pass times, and second-pass times in the same-scene and different-scene conditions.

Axis of rotation	Degree of rotation				Slope in ms/deg
	0	35	70	100	
Same					
X	**Fixation-0 first-pass and second-pass**	220	212	--	--
		947	915	--	0.13
		673	695	--	2.99
Y		227	207	--	--
		887	883	--	-0.33
	272 ms	856	825	--	4.84
Z	906 ms	284	269	--	--
	486 ms	914	945	--	0.55
		532	690	--	2.91
Double		--	--	202	--
		--	--	969	0.63
		--	--	604	1.18
Different					
X	280ms	222	217	--	--
	932ms	935	937	--	0.07
	548ms	407	502	--	0.66
Y		228	200	--	--
		893	912	--	0.29
		516	832	--	3.54
Z		275	277	--	--
		973	1029	--	1.39
		534	783	--	3.36
Double		--	--	206	--
		--	--	982	0.84
		--	--	511	-0.37
Location-2					
X	287ms	216	213	--	--
	978ms	910	970	--	-0.07
	636ms	357	304	--	-4.74
Y		222	196	--	--
		902	840	--	-1.97
		508	648	--	0.14
Z		272	286	--	--
		987	896	--	-1.17
		758	754	--	1.69
Double		--	--	210	--
		--	--	951	-0.27
		--	--	354	-2.82

Continued Next Page

Table 4. Continued

Location-3					
X	310ms	224	217	--	--
	899ms	1074	791	--	-1.54
	440ms	643	540	--	0.14
Y		224	196	--	--
		794	874	--	-0.36
		495	582	--	2.03
Z		309	266	--	--
		998	984	--	1.21
		509	661	--	3.16
Double		--	--	201	--
		--	--	898	0.01
		--	--	444	0.04
Orientation-1					
X	256ms	218	214	--	--
	940ms	916	1089	--	2.13
	593ms	319	802	--	2.99
Y		226	210	--	--
		876	876	--	-0.91
		367	842	--	3.56
Z		296	274	--	--
		1058	1163	--	3.19
		394	790	--	2.81
Double		--	--	210	--
		--	--	1241	3.01
		--	--	906	3.13
Orientation-3					
X	268ms	229	222	--	--
	913ms	840	897	--	-0.23
	522ms	310	363	--	-2.27
Y		240	199	--	--
		979	1060	--	-0.23
		693	1255	--	10.47
Z		251	280	--	--
		849	1075	--	2.31
		475	926	--	5.77
Double		--	--	202	--
		--	--	838	-0.85
		--	--	342	-1.80

Table 5. Response times and error rates in Experiment 2, averaged over all stimulus types.

Axis of rotation	Degree of rotation				Slope in ms/deg and percent error per degree
With Desk	0	35	70	100	
X	**RT and Error**	1882 3.13	2202 4.41	-- --	4.50 0.03
Y		2088 2.76	2299 3.68	-- --	5.89 0.02
Z	1887 ms 2.39 %	2067 3.50	2164 4.05	-- --	3.96 0.02
Double		-- --	-- --	2082 4.78	1.95 0.02
No Desk					
X	**RT and Error**	1904 3.31	2249 7.36	-- --	4.24 0.06
Y		2357 13.24	2530 10.11	-- --	8.26 0.11
Z	1952 ms 3.13 %	1960 3.68	2236 4.23	-- --	4.06 0.02
Double		-- --	-- --	2320 5.70	3.68 0.03

Table 6. Response times and error rates in the same-scene and different-scene conditions in Experiment 2.

Axis of rotation (RT and error rate)	Degree of rotation				Slope in ms/deg and percent error per degree
With Desk	0	35	70	100	
Same					
X	**RT and Error**	1983 1.84	2308 3.68	-- --	6.41 0.05
Y		2222 1.84	2352 2.21	-- --	7.04 0.03
Z	1860 ms 0.37 %	2038 0.74	2152 1.47	-- --	4.17 0.02
Double		-- --	-- --	1954 0.37	0.94 0
Different					
X	1914 ms 4.41 %	1780 4.41	2097 5.15	-- --	2.61 0.01
Y		1954 3.68	2245 5.15	-- --	4.73 0.01
Z		2096 6.25	2177 6.62	-- --	3.76 0.03
Double		-- --	-- --	2209 9.19	2.95 0.05
Location-2					
X	1912 ms 1.47 %	1701 1.47	1990 0	-- --	1.12 -0.02
Y		1097 0	2092 0	-- --	2.58 -0.02
Z		2438 1.47	1748 1.47	-- --	-2.30 0
Double		-- --	-- --	1929 2.94	0.17 0.01
Location-3					
X	1970 ms 0 %	2015 1.47	1733 0	-- --	-3.40 0
Y		1770 0	1833 2.94	-- --	-1.95 0.04
Z		1944 1.47	2466 1.47	-- --	7.09 0.02
Double		-- --	-- --	1841 0	-1.29 0

Continued Next Page

Table 6. Continued

Orientation-1	2038 ms	1917	2713	--	9.65
X	14.71 %	11.76	20.59	--	0.08
Y		1830	2530	--	7.03
		11.76	16.18	--	0.02
Z		2119	2249	--	3.01
		20.59	22.06	--	0.11
Double		--	--	3079	10.4
		--	--	32.35	0.18
Orientation-3	1736 ms	1487	1950	--	3.06
X	1.47 %	2.94	0	--	-0.02
Y		2311	2524	--	11.3
		2.94	1.47	--	0
Z		1880	2244	--	7.27
		1.47	1.47	--	0
Double		--	--	1990	2.54
		--	--	1.47	0
No Desk					
	0	35	70	100	
Same					
X	**RT and Error**	1907	2331	--	6.82
		2.57	6.25	--	0.07
Y		2552	2559	--	10.09
	1853 ms	23.90	14.71	--	0.19
Z	1.10%	1893	2199	--	4.94
		1.10	2.57	--	0.02
Double		--	--	2297	4.44
		--	--	4.04	0.03
Different					
X	2051 ms	1902	2168	--	1.67
	5.15 %	4.04	8.46	--	0.05
Y		2163	2501	--	6.42
		2.57	5.51	--	0.01
Z		2028	2273	--	3.16
		6.25	5.58	--	0.01
Double		--	--	2344	2.93
		--	--	7.35	0.02

Continued Next Page

Table 6. Continued

Location-2	2274 ms	1744	1940	--	-4.77
X	0 %	5.88	0	--	0
Y		2148	2376	--	1.46
		1.47	1.47	--	0.02
Z		2243	1953	--	-4.59
		4.41	1.47	--	0.02
Double		--	--	1917	-3.57
		--	--	2.94	0.03
Location-3	1920 ms	2161	1988	--	0.97
X	1.47 %	5.88	2.94	--	0.02
Y		1919	2224	--	4.35
		1.47	1.47	--	0
Z		1871	2390	--	6.73
		4.41	1.47	--	0
Double		--	--	2038	1.18
		--	--	0	-0.01
Orientation-1	2268 ms	1887	2846	--	8.26
X	19.12 %	4.41	29.41	--	0.15
Y		2006	2352	--	1.19
		4.41	11.76	--	-0.11
Z		2094	2403	--	1.92
		14.71	19.12	--	0
Double		--	--	3154	8.86
		--	--	26.47	0.07
Orientation-3	1743 ms	1815	1899	--	2.23
X	0 %	0	1.47	--	0.02
Y		2578	3050	--	18.68
		2.94	7.35	--	0.11
Z		1905	2344	--	8.59
		1.47	1.47	--	0.02
Double		--	--	2267	5.24
		--	--	0	0

Table 7. Fixation-0 times, first-pass times, and second-pass times in Experiment 2, averaged over all stimulus types.

Axis of rotation	Degree of rotation				Slope in ms/deg
With Desk	0	35	70	100	
X		239	245	--	--
	Fixation-0	826	875	--	0.12
	first-pass	533	733	--	3.04
Y	**and**	250	215	--	--
	second-	840	867	--	0.02
	pass	713	861	--	4.87
Z		293	272	--	--
	278 ms	896	922	--	0.80
	866 ms	633	717	--	2.81
Double	520 ms	--	--	223	--
		--	--	877	0.21
		--	--	635	1.15
No Desk					
X		237	231	--	--
	Fixation-0	816	821	--	-0.46
	first-pass	572	834	--	3.59
Y	**and**	211	186	--	--
	second-	785	859	--	0.08
	pass	1028	1098	--	7.36
Z		276	261	--	--
	278 ms	830	894	--	0.58
	853 ms	612	821	--	3.40
Double	583 ms	--	--	230	--
		--	--	865	0.12
		--	--	822	2.39

Table 8. Fixation-0 times, first-pass times, and second-pass times in the same-scene and different-scene conditions in Experiment 2.

Axis of rotation (RT and error rate)	Degree of rotation				Slope in ms/deg and percent error per degree
With Desk	0	35	70	100	
Same		240	239	--	--
X	**Fixation-0 first-pass and second-pass**	805	859	--	0.02
		631	839	--	4.98
Y		255	218	--	--
		803	858	--	0.02
		845	899	--	5.83
Z		297	267	--	--
	279ms	883	906	--	0.71
	857ms	612	714	--	3.20
Double	490ms	--	--	226	--
		--	--	835	-0.22
		--	--	572	0.82
Different		238	251	--	--
X	277ms	847	891	--	0.22
	875ms	436	627	--	1.12
Y	549ms	246	212	--	--
		878	876	--	0.01
		580	824	--	3.92
Z		289	277	--	--
		909	938	--	0.89
		655	719	--	2.43
Double		--	--	220	--
		--	--	920	0.45
		--	--	698	1.49
Location-2		243	293		--
X	249ms	804	841	--	-0.6
	886ms	421	545	--	0.05
Y	541ms	262	208	--	--
		796	850	--	-0.51
		581	741		2.87
Z		290	266	--	--
		962	822	--	-0.90
		856	492		-0.70
Double		--	--	220	--
		--	--	840	-0.46
		--	--	541	0

Continued Next Page

Table 8. Continued

Location-3		229	226		--
X	328ms	862	842	--	-0.9
	907ms	586	411	--	-1.5
Y	518ms	243	209	--	--
		808	746	--	-2.29
		481	623		1.51
Z		303	279	--	--
		912	899	--	-0.1
		539	955		6.24
Double		--	--	231	--
		--	--	887	-0.20
		--	--	406	-1.12
Orientation-		230	219		--
1	293ms	939	1004	--	2.20
X	850ms	477	1028	--	4.93
Y	683ms	256	219	--	--
		1008	921	--	1.02
		398	1039		5.09
Z		288	275	--	--
		873	1062	--	3.03
		732	665		-0.30
Double		--	--	208	--
		--	--	1053	2.03
		--	--	1287	6.04
Orientation-		251	269		--
3	238ms	782	875	--	0.24
X	858ms	258	526	--	1.01
Y	455ms	224	213	--	--
		901	986	--	1.82
		859	891		6.23
Z		274	289	--	--
		888	967	--	1.55
		492	766		4.44
Double		--	--	222	--
		--	--	900	0.42
		--	--	558	1.03

Continued Next Page

Table 8. Continued

No Desk					
	0	35	70	100	
Same		234	236	--	--
X	**Fixation-0**	790	791	--	-0.73
	first-pass	594	917	--	5.95
Y	**and**	188	183	--	--
	second-	774	858	--	0.22
	pass	1224	1121	--	8.87
Z		302	274	--	--
	277ms	826	867	--	0.36
	842ms	537	804	--	4.35
Double	500ms	--	--	225	--
		--	--	847	0.05
		--	--	805	3.05
Different		240	226	--	--
X	279ms	842	852	--	-0.17
	864ms	549	751	--	1.20
Y	667ms	234	189	--	--
		797	860	--	-0.05
		831	1075	--	5.83
Z		251	248	--	--
		834	921	--	0.83
		688	839	--	2.45
Double		--	--	234	--
		--	--	884	0.2
		--	--	839	1.72
Location-2		268	241	--	--
X	251ms	768	785	--	-0.11
	793ms	454	652	--	-3.92
Y	927ms	248	197	--	--
		784	864	--	1.01
		825	964	--	0.53
Z		242	264	--	--
		907	877	--	1.20
		794	576	--	-5.01
Double		--	--	244	--
		--	--	807	0.14
		--	--	561	-3.65

Continued Next Page

Table 8. Continued

Location-3		234	224	--	--
X	317ms	926	727	--	-1.75
	850ms	723	698	--	2.56
Y	519ms	217	176	--	--
		761	761	--	-1.27
		684	927	--	5.83
Z		276	222	--	--
		848	937	--	1.24
		529	973	--	6.48
Double		--	--	218	--
		--	--	719	-1.31
		--	--	755	2.36
Orientation-1		239	224	--	--
X	302ms	848	1079	--	1.45
	977ms	520	1092	--	5.54
Y	704ms	222	181	--	--
		789	927	--	-0.72
		722	918	--	3.06
Z		253	228	--	--
		751	965	--	-0.17
		818	941	--	3.39
Double		--	--	258	--
		--	--	1146	1.69
		--	--	1202	4.98
Orientation-3		219	216	--	--
X	246ms	826	816	--	-0.26
	834ms	500	562	--	0.63
	518ms				
Y		248	202	--	--
		853	887	--	0.77
		1092	1490	--	13.88
Z		231	276	--	--
		831	907	--	1.04
		611	866	--	4.96
Double		--	--	218	--
		--	--	863	0.29
		--	--	836	3.18

Figure 1. Scene stimuli

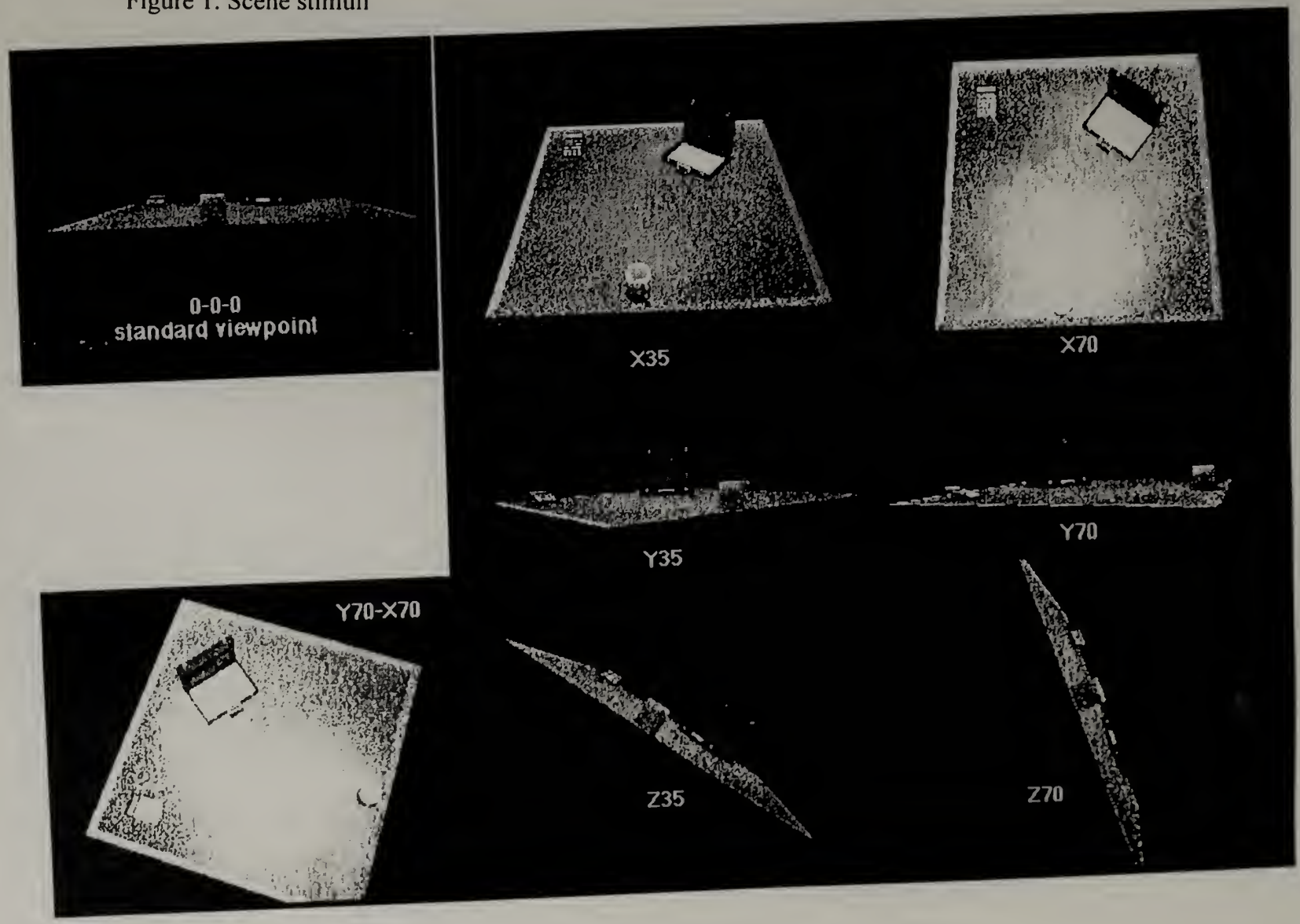
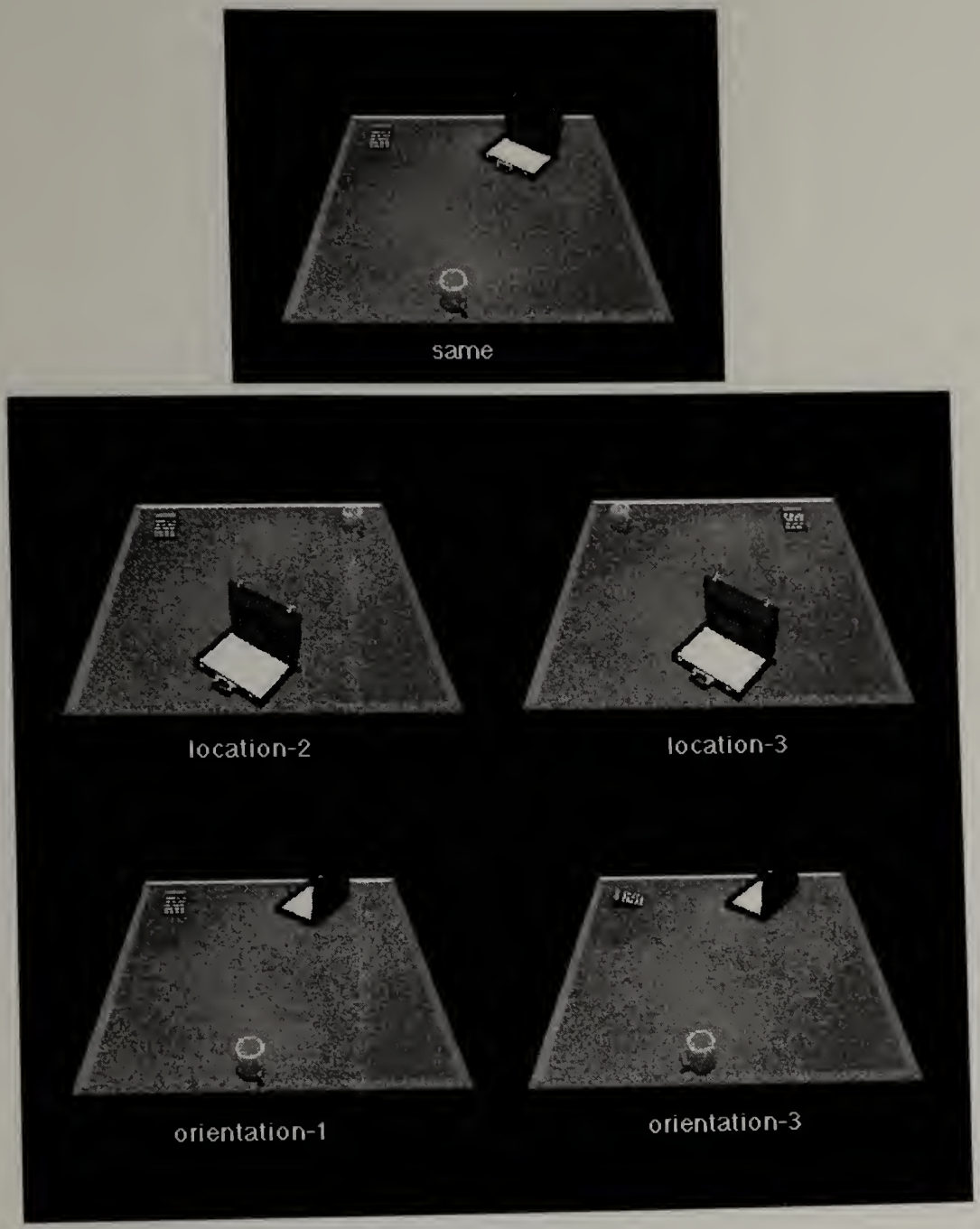


Figure 2. Same-scene and different-scene conditions



BIBLIOGRAPHY

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103*, 62-70.

Antes, J. R., & Penland, J. G. (1981). Picture context effects on eye movement patterns. In D. F. Fisher, R.A. Monty, & J. W. Senders (Eds), *Eye movements: Cognition and visual perception*. Hillsdale: Erlbaum.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 12-23.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1162-1182.

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995) *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1506-1514.

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is richness of our vision an illusion? Transsaccadic memory for complex scenes. *Perception, 24*, 1075-1081.

Boersema, T., Zwaga, H. J. G., & Adams, A. S. (1989). Conspicuity in realistic scenes: an eye-movement measure. *Applied Ergonomics, 20*, 267-273.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Leaning Memory and Cognition, 18*, 531-543.

Cave, K., Pinker, S., Giorgi, L., Thomas, C. E., Heller, L. M., Wolfe, J. M., & Lin, H. (1994). The representation of location in visual images. *Cognitive Psychology, 26*, 1-32.

De Graef, P., de Troy, A., & d'Ydewalle, G. (1992). Local and global contextual constraints on the identification of objects in scene. *Canadian Journal of Psychology, 46*, 489-508.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*, 302-307.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three dimensional objects. *Vision Research, 32*, 2385-2400.

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 395-404.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*, 316-355.

Henderson, J. H. (1992). Visual Attention and eye movement control during reading and picture viewing. In Rayner, K. (Ed.). *Eye movements and visual cognition: Scene perception and reading.* (pp 166-191). New York: Springer-Verlag.

Henderson, J. H. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*, 51-55.

Henderson, J. H., & Anes, M. D. (1994). Roles of object-file review and type priming in visual identification within and across fixations. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 826-839.

Henderson, J., & Hollingworth, A. (1998). Eye movements during scene viewing: an overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception.* Elsevier: North Holland.

Henderson, J., & Hollingworth, A. (1999). The role of fixation position in detecting scene changing across saccades. *Psychological Science, 10*, 438-443.

Henderson, J., Pollatsek, A., & Rayner, K. (1987). Effect of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 449-463.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in neural network for shape recognition. *Psychological Review, 99*, 480-517.

Irwin, D. E., & Carlson-Radvansky, L. A. (1996). Cognitive suppression during saccadic eye movements. *Psychological Science, 7*, 83-88.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory and Cognition, 13*, 289-303.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual difference in spatial ability. *Psychological Review, 92,* 137-172.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565-572.

Logothetis, N., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552-563.

Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology, 4,* 401-414.

Mackworth, N. H., & Morandi, A. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics, 2,* 547-552.

Mannan, S. K., Ruddock, K. H., & Wooding, D. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2D-images. *Spatial Vision, 9,* 363-386.

Mannan, S. K., Ruddock, K. H., & Wooding, D. (1997). Fixation patterns made during brief examination of two dimensional images. *Perception, 26,* 1059-1072.

Nakatani, C., Pollatsek, S., & Johnson S. H. (submitted). Viewpoint dependent recognition of scenes.

Nelson, W.W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 391-399.

Parsons, L. M. (1987). Visual discrimination of abstract mirror reflected three-dimensional objects at many orientations. *Perception and Psychophysics, 42,* 49-59.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalization of recognition without mental transformations. *Cognition, 67,* 111-145.

Pollatsek, A., & Rayner, K. (1992). What is integrated across fixations? In Rayner, K. (Ed.). *Eye movements and visual cognition: Scene perception and reading.* New York: Springer-Verlag, 166-191.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 112,* 426-442.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263-266.

Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multiplayer network. *Sceince, 247*, 987- 982.

Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neurosceience, 2*, 1019-1025.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Experimental Psychology, 46*, 342-376.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for a parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the macrostructure in cognition 1.*

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects, *Science, 191*, 952-954.

Tarr, M. J. (1995). Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2*, 55-82.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*, 1-20.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*, 233-283.

Van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual of the visual field: moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception.* Elsevier: North Holland.

Wachsmuth, E, Oram, M. W., & Perrett, D. I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex, 4*, 509-522.

Wurtz, R. H., & Munoz, D. (1994). Role of monkey superior colliculus in control of saccades and fixation. In M. Gazzaniga (Ed.), *The cognitive neurosciences.* Cambridge, MA: MIT Press.